Synthetic Biology Toolkit for *Cupriavidus necator* H16: An Industrially Relevant Microbe

Christopher Chibueze Azubuike

A thesis submitted for the degree of

Doctor of Philosophy in Biology

School of Natural and Environmental Sciences (SNES)
Newcastle University, Newcastle upon Tyne
United Kingdom

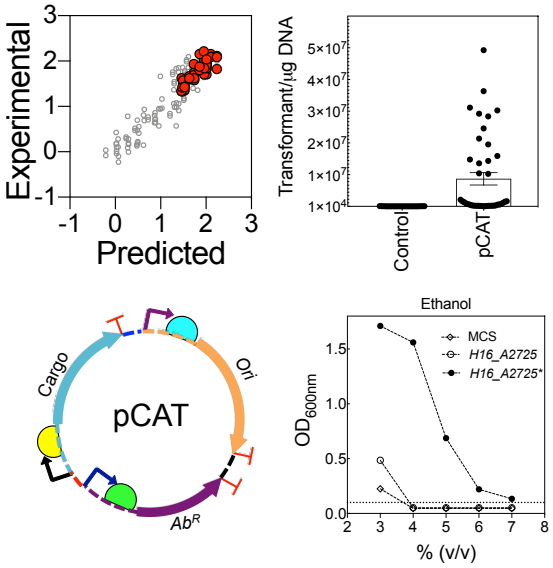
May 2020

Declaration

I declare that the work presented in this thesis is conducted solely by me, and it has not been submitted elsewhere for the award of any degree. Sources of information mentioned in this work are rightfully acknowledged.

Christopher C. Azubuike

Abstract



Cupriavidus necator is a Gram-negative soil bacterium of great biotechnological interest. It is known as a producer of the bioplastic 3-polyhydroxybutyrate, has been used in bioremediation efforts, and its lithoautotrophic capabilities raise the possibility that it could function as a microbial factory upgrading renewable resources to fuels and chemicals. However, appropriate experimental resources to permit controlled bioengineering and system optimisations with the bacterium are not well established. In this study, statistical Design of Experiments (DoE) was used to identify how key media components and their interactions affect cell growth. The model resulting from this approach is predictive and was experimentally validated against novel media compositions at different cultivations scales. Specifically, interaction between histidine and $CuSO_4$ are important for reliable robust growth prediction. Further, plasmid parts (replication origins, antibiotic cassettes and reporter proteins) were characterised for improved transformation efficiency, segregational stability and reporter protein expression in *C. necator*. Modular minimal plasmid sets (pCAT) were constructed for *C. necator* using these well-characterised biological parts, which were assembled by Golden gate method. The resulting plasmids were delivered to *C. necator* via electroporation, and the transformation efficiency obtained was more than 3000-fold higher in comparison to that obtained with the existing plasmid, pBHR1. More importantly, the resulting Golden gate restriction-ligation products can be delivered directly to *C. necator* via electroporation with high transformation efficiency and can co-express more than one functional protein carried on a single plasmid restriction-ligation product. pCAT plasmids can stably propagate for more than six generation (144 h) without the addition of antibiotics. Furthermore, the application of the toolkit was demonstrated by engineering *C. necator* for improved tolerance to ethanol using directed evolutionary approach. The toolkit established in this study will be crucial in making future bioengineering applications in *C. necator* more efficient, controllable and predictable.

Keywords: *Cupriavidus necator* | Design of Experiments (DoE) | Defined media | Plasmid | Electroporation | Mutagenesis

Acknowledgements

My sincere gratitude to the Commonwealth Scholarship Commission for full PhD scholarship (NGCA-2016-60) and to the University of Port Harcourt Nigeria for this great opportunity.

My PhD journey is smooth because of the professional advice from the Principal investigator of my research, Dr. Thomas P. Howard, and also from the co-investigator, Prof. Angharad M. R. Gatehouse. I am very grateful to all your support and expert advice. I am also very thankful to Dr. Martin G. Edwards for creating the opportunity that enabled me to explore in other areas of professional development.

To all the staff and colleagues in the 5th Floor Devonshire building and lab 4.7, I am very thankful to your kindness and support. Specifically, to my progression panel Drs.: James Stach and Maria C. Montero-Calasanz for their constructive feedback throughout my progression. My sincere thanks to Dr. Mathew Peake, Dr. Samuel Logan and Gillian Davison for their assistance in procuring the consumables used throughout my research. Additionally, to the Howard research group: Dr. Colette J. Whitfield, Dr. Alice M. Banks, Joshua Loh, Alex Laverick and Alis Prusokas for their support.

I highly appreciate Drs.: Chioma B. Chikere, Jennifer C. Nwosu and Nwamaka Okeke-Ogbuafor for their good counsel. And to my good friends: Chukwuebuka P. Opara, ThankGod P. Nwogu, Peter C. Mbiaka, and Lilian C. Izuegbulam for their love and care.

Lastly to my family: Chief Mr. and Mrs. Christopher E. Azubuike; Athanasius, Michael, Patrick, Augustine and Anastasia; Joseph, Emmanuel, MaryAnne; Jennifer and Chisom for supporting me.

Table of Content

Contents

List of Tables

List of Figures

Abbreviations

Ab^R	Antibiotic resistant cassette
BHR	Broad host range
CBB	Calvin-Benson-Bassham
CDM	Chemically defined medium
CFU	Colony forming unit
CmR	Chloramphenicol resistant cassette
CRISPR	Clustered regularly interspaced short palindromic repeats
dO_2	Dissolved oxygen
DoE	Design of Experiments
DSD	Definitive Screening Design
ED	Entner-Doudoroff
EFE	Ethylene forming enzyme
eGFP	Enhanced green fluorescent protein
Goi	Gene of interest
gTME	Global transcription machinery engineering
IPTG	Isopropyl β-d-1-thiogalactopyranoside
KanR	Kanamycin resistant cassette
KDPG	2-keto-3-deoxy-6-phosphogluconate
LB	Lysogeny broth
M-H	Mueller-Hinton
MBC	Minimal bactericidal concentration
MCS	Multiple cloning site
MIC	Minimal inhibitory concentration
MoA	Mode of action
Mob	Mobilisation region
mRFP	Monomeric red fluorescent protein
NHR	Narrow host range
OD	Optical density
OFAT	One factor at a time
Ori	Origin of replication
OriT	Origin of transfer
PBAT	Polybutyrate adipate terephthalate
PCR	Polymerase chain reaction
PDCA	Pyridinedicarboxylic acid
PHA	Polyhydroxyalkanoate

PHB	Polyhydroxybutyrate
Rep	Replication protein
RFU	Relative fluorescent unit
SEM	Standard error mean
SEVA	Standard European Vector Architecture
SmR	Streptomycin resistant cassette
SOB	Super optimal broth
SOC	Super optimal broth with catabolite repression
TEM	Transmission electron microscopy
TcR	Tetracycline resistant cassette
TCA	Tricarboxylic acid cycle
vvm	Volume of air per volume of medium

Chapter 1.

Progress on the Development of Genetic Toolkit for *Cupriavidus necator* H16 – a Bioplastic Producing Bacterium

1.1 Introduction

Non-model microbes are gaining significant attention as promising chassis for future biotechnological applications. Their ability to carry out complex biological processes makes them more favourable over *Escherichia coli* and *Saccharomyces cerevisiae,* the dominant industrial biocatalysts [1,2]. Genome engineering tools for these two groups of microbes are by far more advanced than for any other microbe [3,4]. *E. coli* dominates the synthetic biology field due to the availability of vast amount of advanced and well-characterised tools for manipulating the bacterium. Whilst its short doubling time (20 min) facilitates rapid bioengineering studies. On the other hand, yeasts, including model and non-model yeasts, have more tolerance to sugars and can utilise range of compounds including aromatics [5]. Synthetic biology as a multidisciplinary field is advancing from laboratory demonstrations to practical applications requiring robust chassis organisms to meet the needs of the society. One of these needs is the ability to harness renewable resources to produce high-valued compounds; another is the ability to be deployed as a bioremediation agent. *E. coli* and yeasts are not genetically equipped to simultaneously meet these two needs. In harnessing renewable resources, chemolithoautotrophic bacteria offers much. Their ability to fix CO_2 offers a platform for utilising 'greenhouse gas' as a source of carbon for biosynthesis [6]. It is very uncommon to find an organism that is genetically equipped to serve the dual purpose of biosynthesis and bioremediation. Fortunately, *Cupriavidus necator* H16 a 'Knallgas' bacterium is an excellent fit for this dual purpose. The bacterium is non-pathogenic thus generally recognised as a safe.

Cupriavidus necator H16 (hereafter *C. necator*, unless mentioned to distinguish between isolates) is a model chemolithoautotrophic *β*-proteobacterium well-known for its ability to accumulate > 80% of its dry cell weight as poly(*R*)-3-hydroxybutyrate (PHB)—a biodegradable polymer—under carbon-rich and macronutrient(s) limiting growth conditions [7,8] (Fig. 1.1A). While the bacterium can grow on range of organic carbon substrates including alcohols, amino acids, fatty acids and sugar acids [9], in the absence of organic carbon, it can also grow robustly by fixing CO_2 via the Calvin-Benson-Bassham (CBB) cycle, using H_2 or formate as an energy source (Fig. 1.1B). The bacterium is also capable of using

alternative electron acceptors NO_3^- or NO_2^- to respire by denitrification in the absence of oxygen [10]. These metabolic features mean that *C. necator* has garnered much interest as a production platform for biofuels, biopolymers and other commodity chemicals. Industrial biotechnological applications of *C. necator* have been demonstrated through the microbial synthesis of higher alcohols [11–15], fatty acids [16–18], alkanes [19], polymers and polymer precursors [20,21], enzymes [22] and other high-valued compounds [23] both under heterotrophic and autotrophic growth conditions. Furthermore, the of use carbon monoxide as a carbon source to accumulate biopolymer was demonstrated for recombinant strain of *C. necator* [24]. Compared to other autotrophic bacteria, *C. necator* has a faster growth rate [19]. These features make *C. necator* a competitive robust microbial chassis. Other applications of *C. necator* have been extensively reviewed [8,23].

Although *C. necator* fulfils the criteria as a robust next generation microbial chassis, like other non-model microbial chassis, lack of knowledge on existing genetic tools delays the realisation of the full potentials of the bacterium. Extensive reviews on other non-model bacterial chassis have been carried out [2,6]. Therefore, this review focuses on tool development for this industrially relevant biocatalyst. It provides an update on toolkit for *C. necator*, limitations of some of the tools and offered insight on how the limitations can be solved. Such information will help advance research beyond what is currently explored with the bacterium specifically in biosynthesis, biotransformation and bioremediation.

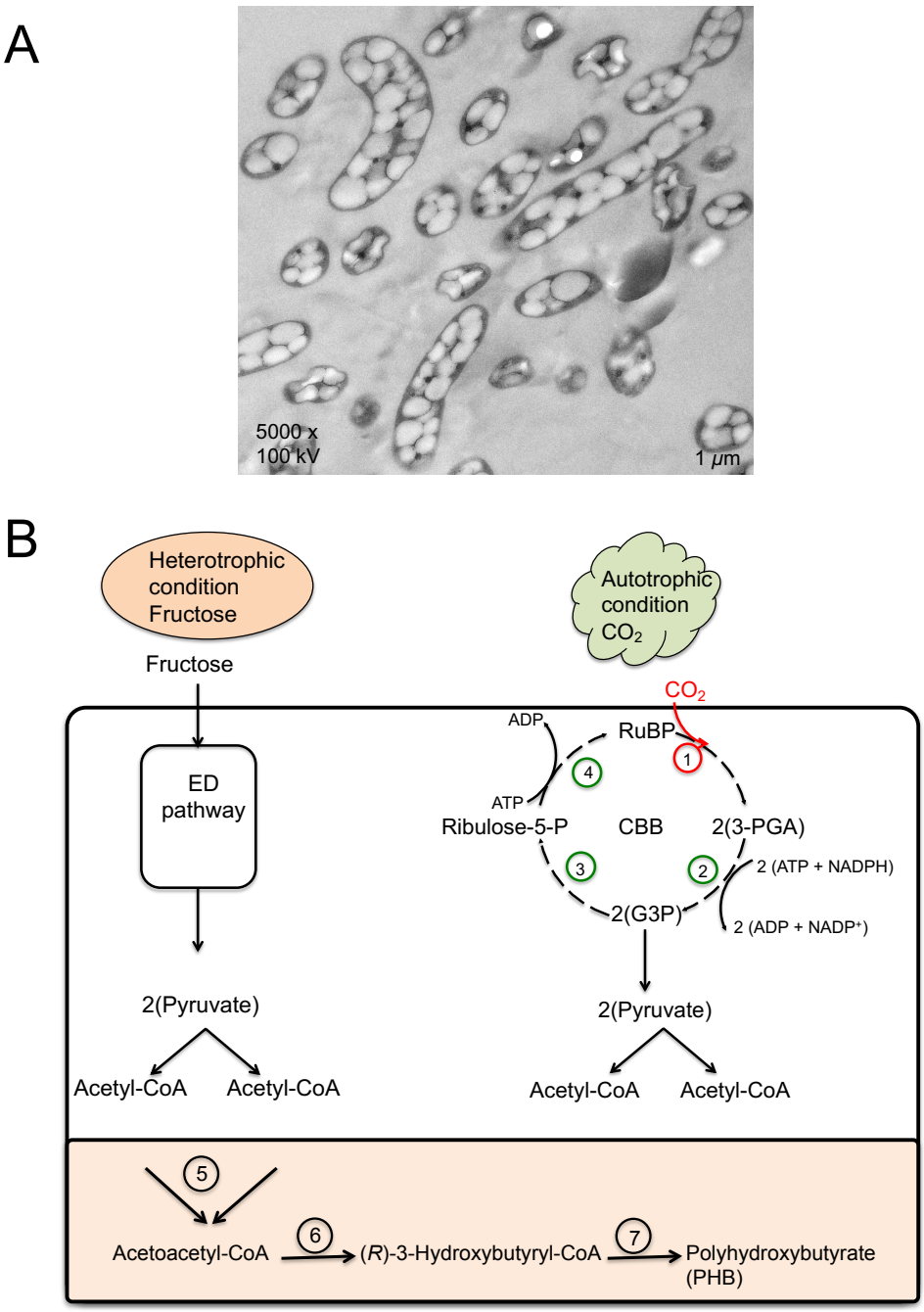


Fig. 1. 1 *C. necator* notable metabolic feature.

A. PHB accumulating *C. necator* observed under TEM. **B.** Scheme of *C. necator* central metabolic pathways leading to accumulation of polyhydroxybutyrate (PHB) under heterotrophic or autotrophic condition. Fructose is imported via ABC transporter and catabolised via the Entner-Doudoroff (ED) pathway; 2-keto-3-deoxy-6-phosphogluconate (KDPG) is a key intermediate of the pathway. Under lithoautotrophic condition CO_2 is fixed in the first phase of Calvin-Benson-Bassham (CBB) cycle into a stable intermediate, which is further reduced in the second phase to a carbohydrate. The numbers indicate enzymes involved: **1**. RuBisCo catalyses carboxylation of ribulose bisphosphate (RuBP); the resulting intermediate gives rise to 2 molecules of 3-phosphoglycerate (3-PGA); **2**. Glyceraldeyhde-3-phosphate dehydrogenase reduces 1,3-bisphosphoglycerate (from 3-PGA phosphorylation by phosphokinase) to glyceraldehyde-3-phosphate (G3P); **3**. Phosphopentose isomerase converts ribose-5-phosphate (a product of regeneration phase) to ribulose-5-phosphate (Ru5P); **4**. Phosphoribulokinase phosphorylates Ru5P to RuBP to complete the CBB cycle. The broken cycles (stages) indicate further reactions occurring before the ones depicted. The key product of both ED pathway and CBB cycle are 2 pyruvate molecules, which is further decarboxylated to acetyl-CoA. **5**. β-ketothiolase under growth limiting condition condenses two molecules of acetyl-CoA to acetoacetyl-CoA. **6**. Acetoacetyl reductase reduces acetoacetyl-CoA to *R*-3-hydroxybutyryl-CoA (HB). **7**. A synthase polymerises (*R*)-3-hydroxybutyrate (HB) to polyhydroxybutyrate (PHB).

1.2 Isolation and nomenclature

The nomenclature of this industrially relevant bacterium is fascinating and provides a clear insight into its metabolic capabilities. Its isolation can be traced to few years before 1962. Three strains (H1, H16 and H20) were isolated from sludge sample and were classified under the genus *Hydrogenomonas*, due to their ability to grow with oxyhydrogen and CO_2 in a mineral medium [25]. *Hydrogenomonas* H20 had the shortest doubling time, 195 min, compared to isolates H1 and H16 with doubling time 220 and 270 min, respectively. Although an unpublished study [26] stated *Hydrogenomonas eutropha* was first isolated in 1957 by Bovell who added the species name *eutropha*—without detailed description of the bacterium—both studies agreed on the ability of the bacterium to grow autotrophically with H_2 as the energy source. In 1969, a proposal to reject the genus *Hydrogenomoas* due to isolation of other bacteria with similar ability to grow autotrophically with H_2 but differ significantly in other morphological and nutritional properties was successful [27]. In this new system, *Hydrogenomonas eutropha,* Bovell strain, was assigned *Alcaligenes eutropha*. This was because of its general, as well as distinguishable properties (Fig. 1.2). *Alcaligenes eutropha* is a rod shaped, non-pigmented bacterium with peritrichous flagella. It can grow on organic carbon sources and has high percentage (66.3–66) G + C content. Nevertheless, it was suggested that the genus *Alcaligenes* might be inappropriate for the bacterium [27]. Further, it was made clear that *A. eutropha* and *Hydrogenomonas* H20 are the same strain. Twenty-six years later (in 1995), based on phenotypic features, cellular compositions and phylogenetic analysis, *A. eutropha* was transferred to *Ralstonia eutropha*—a new genus named after an American Bacteriologist, E. Ralston [28]. However, in 2004, due to isolations of new species belonging to *Ralstonia*, and to distinguish between the two lineages of the genus: *R. eutropha* and *R. pickettii*, a new genus was proposed for *R. eutropha* [29]. This new genus, *Wautersia*, was assigned to *R. eutropha* lineage owing to their peritrichous flagella, inability to produce acid from glucose fermentation, susceptibility to colistin, and longer viability on tryptic soy agar at 25°C in comparison to their counterpart, *R. pickettii*. Named after a Belgian Microbiologist (George Wauters), *W. eutropha* (DSM 531) was further distinguished from other *Wautersia* species due to its ability to assimilate L-serine, *N*-acetylglucosamine and 2-ketogluconate, and inability to alkalinise mucate on Simmon's agar base. This new genus, *Wautersia*, only lasted few months. In the same year *Wautersia* species were renamed *Cupriavidus*, an already existing genus [30]. The reclassification was in accordance with the International Code of Nomenclature of Bacterium (ICNB), on the basis that *Cupriavidus necator* type species (LMG 8453^T = ATCC 43291), with 79% DNA-DNA hybridisation and 99.7% 16S rRNA sequence similarity to *Wautersia eutropha* (LMG 1199^T = ATCC 17697), had first been described 17 years earlier (in 1987) [31].

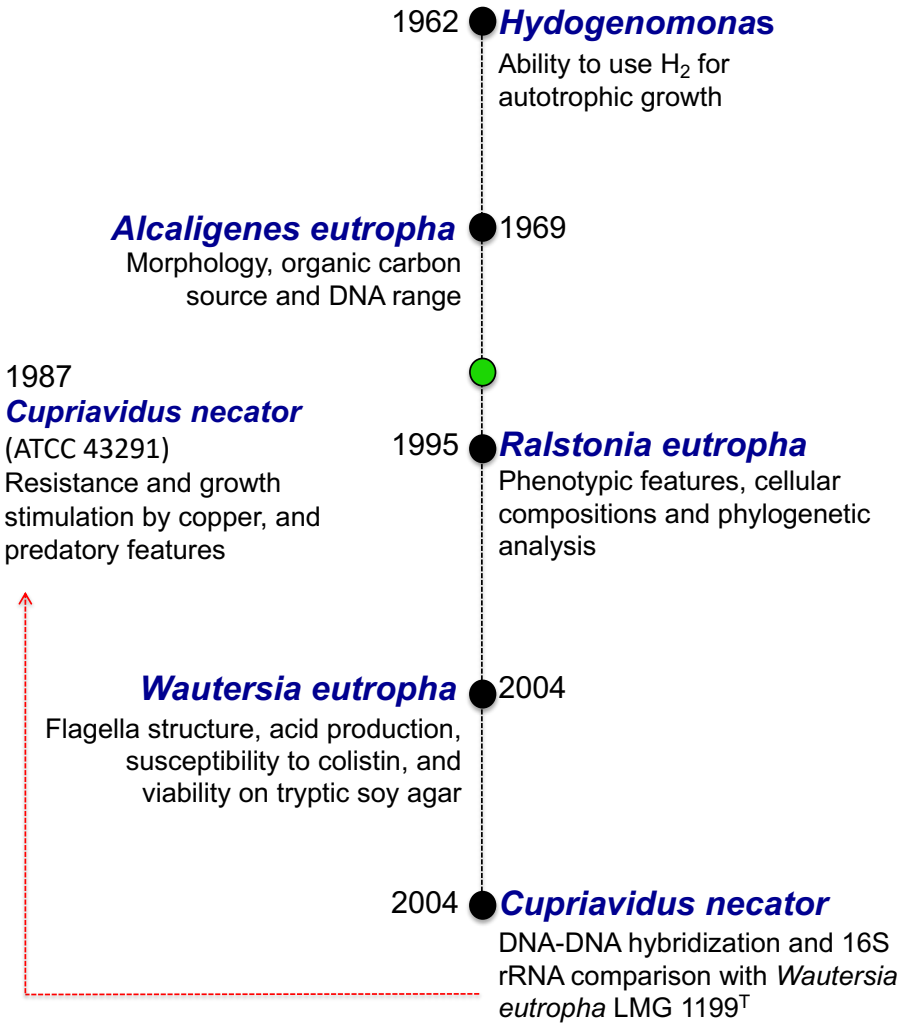


Fig. 1. 2 Timeline in *C. necator* nomenclature.

Unique feature(s) of the bacterium, for which it was renamed at a given point in time, is giving below the genus and species name. The green circle indicates when the name *C. necator* was first used to describe a different type species of *C. necator* other than H20 or H16, with the red dotted line tracing the current accepted name back to when it was first used. The type species used throughout the reclassification was strain H20 (ATCC 17697^T, = CCUG^T = DSM 532^T = LMG 1199^T).

Having understood the nomenclature of the genus *Cupriavidus*, it can be deduced that *C. necator* H16 was *Hydrogenomas* H16 (ATCC 17699 = LMG 1201), which was reported by Wilde [25] as having the lowest doubling time in comparison to strains H1 (DSM 529) and H20. Earlier study of PHB accumulation using organic substrates as the carbon source was carried out using strain H16 [25]. This study observed that PHB was the only storage compound formed. The hydrogenase from *C. necator* H16 is well regulated and can be synthesised in the presence of H_2 and under poor growth condition [32].

With knowledge of its nomenclature, settled for now, *C. necator* H16 can be described as a Gram-negative, rod-shaped, non-spore forming, non-pigmented, motile bacterium, with ability to grow aerobically and anaerobically, and able to grow chemolithoautotrophically using H_2 as energy source. It has been isolated from soil and clinical samples. In addition, the bacterium is known as a non-obligate predator (slayer, which in Latin means *necator*) of bacteria, and a copper lover, hence the name *Cupriavidus necator*.

1.3 Genome and proteome

The microbiological features of *C. necator* H16 were fundamental in its early industrial applications. However, it was not until 2006, following the complete genome sequence of the bacterium's chromosomes, that its industrial biotechnological applications witnessed a surge. The genome of the bacterium comprise three replicons: chromosome 1, chromosome 2 and a mega plasmid (pHG1), totaling 7.4 Mbp [9]. Prior to the genome sequencing, pHG1 was the first replicon to be mapped and fully sequenced [33,34]. Genes responsible for the flexible bioenergetic features of the bacterium are distributed across the replicons (Fig. 1.3). Chromosome 1 encodes genes for DNA replication, key players of aerobic and organic acid metabolism and *R*-3-polyhydroxybutyrate (PHB) synthesis. Chromosome 2 encodes genes for fructose and sugar acid metabolism, whilst that for hydrogen oxidation is encoded on the plasmid, pHG1. Genes for aromatic compound degradation, anaerobic respiration and CO_2 fixation (*cbb* genes) are distributed on chromosome 2 and pHG1. Formate dehydrogenase genes are located on chromosome 1 and 2. It appears that chromosome 2—with plasmid-like origin—encodes the majority of the flexible growth features, which it shares with chromosome 1 and pHG1 (Fig. 1.3). Given that the genes for autotrophic growth and aromatic compound degradation are borne on both chromosome 2 and pHG1, it can be surmised that the bacterium originally was not equipped to grow under autotrophic condition or able to tolerate pollutants. The ability to carry out autotrophic metabolism, which appears to be an adaption to limited resources (energy and carbon), might have been acquired by horizontal or lateral gene transfer via plasmid; the genes then became transposed, recombined, or integrated into the chromosome. This is supported by the comparative genomic of *C. necator* isolates [35]. Chromosome 2 contains copies of rRNA indicating past recombination event. *C. necator* H16 is naturally a heterotroph, which maintains heterotrophic metabolism (active tricarboxylic acid cycle) even under autotrophic condition [36]. Further, when cured of pHG1 *C. necator* H16 was able to retain its heterotrophic metabolism but lost lithoautotrophic capability [33,37].

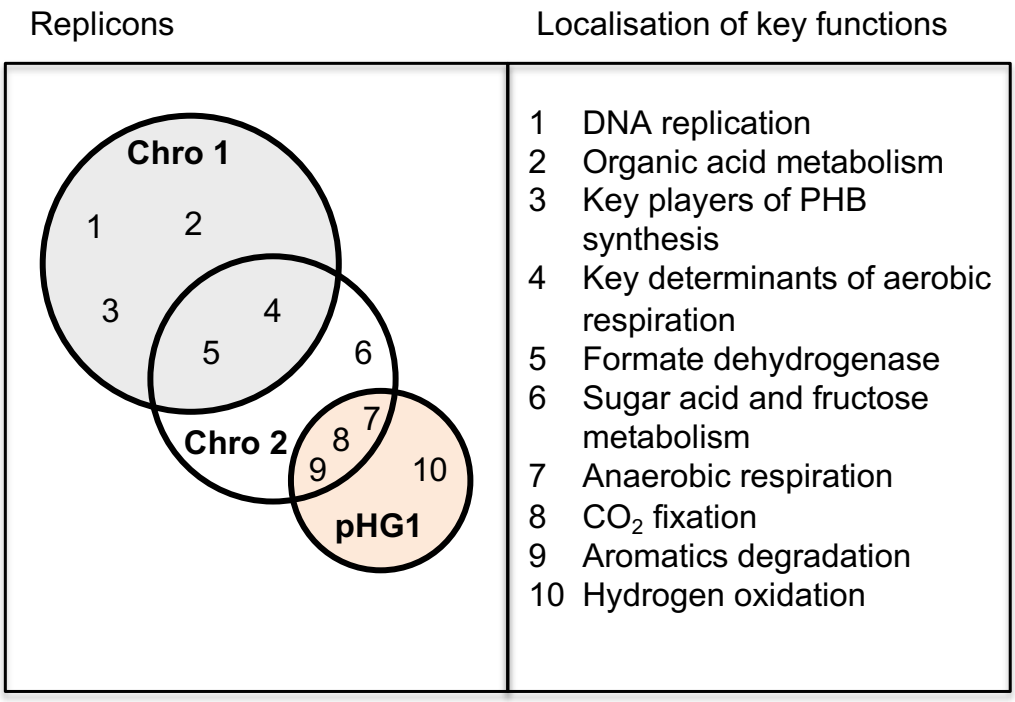


Fig. 1. 3 Localisation of key genes responsible for *C. necator* H16 metabolic features.

The genome of the bacterium consists of three replicons: chromosome 1 (chro 1), chromosome 2 (chro 2) and a mega plasmid (pHG1). Key genes coding for the different metabolic functions are arbitrary depicted.

Analyses of *C. necator* H16 proteome revealed differential expression of proteins under different growth conditions [36,38]. The majority of the proteins identified were encoded on chromosome 1 in comparison to chromosome 2 and pHG1. Although the observation was ascribed to chromosome 1 encoding genes for central metabolism and DNA replication, it is possible that such underrepresentation of proteins from the other two replicons might be due to the growth conditions tested [36]. Similarly, quantitative proteome analysis revealed overrepresentation of proteins whose genes are located on chromosome 1 [38]. More soluble proteins were detected compared to membrane proteins. More protein fractions were detected for chromosome 2 and pHG1 under lithoautotrophic condition in comparison to those detected under heterotrophic condition thus demonstrating that these two replicons coordinate most of the metabolic activities under lithoautotrophic growth condition (Fig. 1.3). In addition to growth conditions, i.e. cultivation under autotrophic or heterotrophic mode, growth media and carbon sources might have contributed to the observed differential protein expression across the three replicons. More quantitative proteomic analyses especially under anaerobic condition will shade light into the proteins that are expressed under such condition. Regulation of such proteins will improve the scope of *C. necator* H16 biotechnological applications.

1.4 Growth media and cultivation conditions

One of the early distinguishable features of *C. necator* is its nutritional requirement. This was extensively studied for the type strain, *C. necator* H20, under autotrophic condition [26]. The bacterium is reported to have a doubling time of 195 mins. The optimum growth temperature observed after 24 h cultivation was between 25–30°C [26,31]. Above 30°C growth was impaired. In a medium with buffering agent, the optimum pH, for which maximum growth rate was recorded within 12 h cultivation was 6.4–6.9 [26]. Nevertheless, *C. necator* has a reported broad pH tolerance range, 5.5–9.2 [31]. The effect of pH on *C. necator* growth is more on growth rate than on the overall growth over a period of time. Under chemolithoautotrophic condition, 15–25% oxygen concentration was considered optimum, and within this range similar growth rates were obtained [26]. Oxygen concentration (less than 10%) impacted negatively on growth [26]. Under optimum oxygen concentration, 15% CO_2 supported robust growth. However, growth was impaired by up to 50% when the concentration of CO_2 was reduced to 5% at the same optimum (20%) oxygen concentration. Furthermore, amongst the nitrogen sources tested only nitrite failed to support growth. Nickel and iron are essential for cultivation of *C. necator* H1 and H16 under autotrophic condition [39,40]. These observed growth requirements under autotrophic condition guided cultivations under heterotrophic condition especially for *C. necator* H16.

Under heterotrophic condition *C. necator* are able to utilise organic acids and few sugars as carbon sources [25,31,41]. Specifically, sugar metabolism by *C. necator* H16 is limited to fructose and *N*-acetylglucosamine [9]. Fructose is metabolised via the Entner-Doudoroff pathway (Fig. 1.1), with the responsible genes located on chromosome 2 (Fig. 1.3), whilst the genes for *N*-acetylglucosamine metabolism are located on chromosome 1 [9]. The inability of the wild type strain to metabolise glucose is linked to the absence of phosphofructokinase and 6-phosphogluconate dehydrogenase—key enzymes of the Embden-Meyerhoff-Parnas and oxidative pentose pathways, respectively [9]. Together, these findings formed the basis of defined media formulations and growth conditions for *C. necator* cultivations [12,13,16,20,42–45]. In this study, a statistical Design of Experiments (DoE) was employed to study the impact of media components on *C. necator* growth under heterotrophic condition Chapter 2. This systematic approach highlighted media components and interactions that contribute significantly to the growth of *C. necator*. As expected, fructose supported the best growth compared to the other carbon sources tested: glucose, glycerol and sucrose. Positive interaction was observed between amino acid and trace elements, with copper and histidine the major contributors to the interaction. Although the bacterium is capable of growing in media without amino acids, addition of a few amino acids—at a moderate concentration—in

media increased *C. necator* growth rate. The components that were found to be essential for *C. necator* growth were only fructose and magnesium, whilst potassium and amino acids were considered important. The bacterium prefers amino acids to ammonium as a nitrogen source to begin its growth. A good understanding of how *C. necator* responds to basic media components and cultivation conditions is imperative for exploiting the industrial potentials of the bacterium.

1.5 Plasmid

A prerequisite for a microbial chassis is the ability for efficient transformation with foreign genetic materials. Typically, this is achieved through the use of plasmids. Plasmids are made of three important components: an origin of replication, antibiotic resistant gene and a cargo (gene of interest (goi) or multiple cloning site for insertion of goi)) (Fig. 1.4A). These three components are critical in bioengineering application of *C. necato*r. The bacterium can be transformed with a range of foreign genetic materials carried on a plasmid. However, the plasmid range that is propagated in *C. necator* is narrow due to some plasmid parts (components) being unstable or incompatible in the bacterium. Hence, information on the compatibility of plasmid part is vital for genetic manipulation of this industrial relevant bacterium.

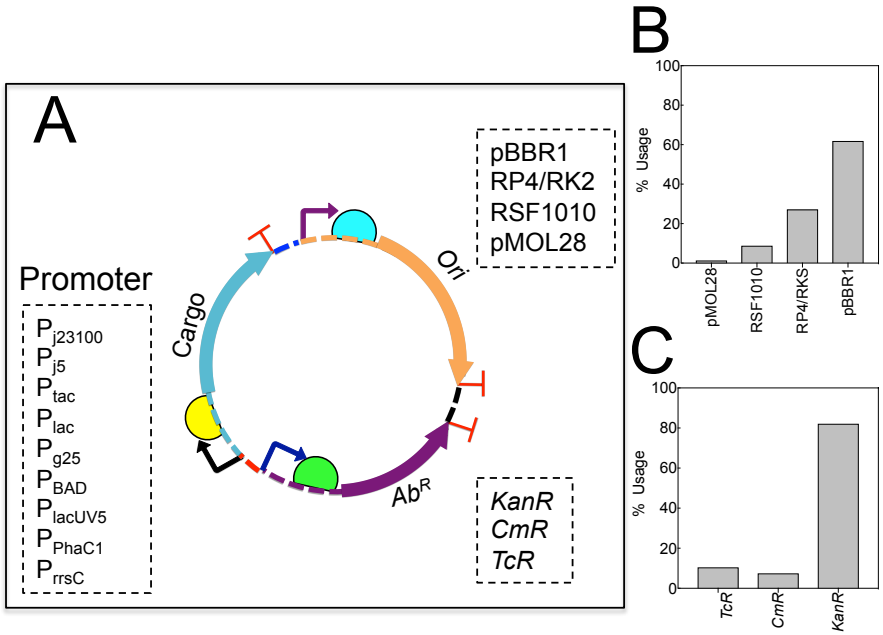


Fig. 1. 4 Plasmid for *C. necator* application.

A. Illustration of a modular plasmid backbone for *C. necator* showing the major biological parts (bioparts). Origin of replication (*ori*), antibiotic resistant cassette (Ab^R) and cargo (goi or MCS) form the major biopart. Each biopart is preceded by a promoter (curved arrow) driving gene expression. Located between the promoter and gene is a ribosomal binding site (coloured chord). At the end of each gene is a terminator (T). Ab^R: kanamycin (*KanR*); chloramphenicol (*CmR*); tetracycline (*TcR*). **B**. Percentage use of origin of replication. **C**. Percentage use of antibiotic resistant cassette. Data for origin of replication and antibiotic resistant cassette were from more than 80 genetic related studies involving *C. necator* H16 as the host from 1981 to 2019.

1.5.1 Plasmid origin of replication

An origin of replication (*oriV*) is essential for autonomous replication of a plasmid in a host. In most cases, *oriV* and its initiator protein(s) determine plasmid compatibility in a given host in order to initiate and regulate plasmid replication. Plasmids belonging to different incompatibility groups have been successfully propagated in *C. necator* [46]. Among them, plasmids bearing pBBR1 *oriV* account for more than 60% of plasmids that are used in transforming *C. necator* (Fig. 1.4B). pBBR1, with an undefined incompatibility group, are widely used for studies involving *C. necator*. This is likely because pBBR1 is relatively small in size compared to pMOL28, RP4/RK2, and RSF1010 [47]. And unlike RSF1010, which requires a primase (Rep B), a helicase (Rep A) and an initiation protein (Rep C) located distant from its *oriV* for autonomous replication, pBBR1 *oriV* requires single replication protein, Rep—encoded in its putative region—to initiate replication [46].

There are many variants of plasmids bearing pBBR1: pBBR1MCS-1–pBBR1MCS-5, and pBHR1 [48,49]. These plasmids bear multiple cloning sites (MCS) that facilitate cloning, and different antibiotic resistant cassettes. Further, pBBR1 is a medium copy number replicon. And like RP4 and RSF1010, it is stably maintained in several Gram-negative bacteria including *C. necator*. It appears that narrow host range *oriV* plasmids (ColE1, pF1, p15A, pSC101, pMB1 and its derivative, pUC19) are not able to direct autonomous replication in *C. necator* Chapter 4; however, they lend themselves in the construction of suicide vectors for gene replacement via homologous recombination in *C. necator*. Interestingly, pBBR1 has been co-propagated with IncP and IncQ plasmids [50]. This means that pBBR1 does not belong to either of the incompatibility groups. More studies are needed to assign an incompatibility group to pBBR1. The lack of knowledge of pBBR1 incompatibility has impeded some aspects of cloning in *C. necator*. Specifically, plasmid variants of pBBR1 with very high yields in *E. coli* following purification are not well propagated in *C. necator* [51], Chapter 4. Conversely, similar variants with low yields are well propagated and established under prolonged cultivation. In addition, pBBR1 is believed to have both *oriV* and *Rep*, which perform different function to ensure plasmid replication across different bacteria hosts. Intriguingly, both genes were found to independently direct plasmid replication across bacteria Chapter 4. Nonetheless, plasmid replication and segregational stability are significantly better when both genes are involved. Thus, understanding pBBR1 incompatibility will contribute significantly to improving bioengineering applications in *C. necator* with this plasmid group.

1.5.2 Antibiotic selection marker

Another essential part of a plasmid is antibiotic resistant gene. The resistance gene encoded by naturally occurring or synthetic plasmids also determines the host range of a given plasmid. Some group of bacteria, due to their cell envelope, are prone to be more susceptible or resistance to some group of antibiotics. For example, Gram-positive bacteria are more susceptible to antibiotics that inhibit cell wall synthesis. This is due to Gram-positive bacteria having more peptidoglycan layer compared to Gram-negative bacteria. Thus, plasmids coding for resistance against broad-spectrum antibiotics are more likely to be acquired by distantly related groups of bacteria in an environment via any means of genetic transfer. Generally, plasmids employed in *C. necator* studies encode mostly chloramphenicol, kanamycin or tetracycline resistant genes resulting in the use of these antibiotics for selecting *C. necator* transformants [50,52–55]. Approximately 80% of plasmids used in *C. necator* studies bear kanamycin resistant gene (Fig. 1.4C)

Antimicrobial susceptibility of wild type *C. necator* H16 revealed an interesting susceptibility pattern [Chapter 3]. The bacterium is sensitive to chloramphenicol, kanamycin, carbenicillin, and showed high degree of sensitive to tetracycline even at very low concentration, < 1 μg/mL. This antibiotic, tetracycline, is considered as having one of the best broad spectra of activity, effectively inhibiting microbial protein synthesis both in Gram-positive and Gram-negative bacteria. Additionally, the bacterium displayed inconsistent susceptibility to ampicillin. False positive colonies were observed on agar plate supplemented with erythromycin following transformation of *C. necator* with plasmid bearing erythromycin resistance gene. Thus, ampicillin and erythromycin are not considered suitable for selecting *C. necator* transformants. The bacterium is moderately sensitivity to spectinomycin and has been designated as gentamicin resistance [12,44,56].

1.5.3 Promoter

An important region in a given goi is the upstream region before the translational start site, which determines the activity and level of gene expression. This region is called a promoter, and it comprises of different genetic elements, which together determine the strength of a given promoter [57]. Different sets of native, synthetic, constitutive and inducible promoters have been studied for applications in *C. necator* (Table 1.1).

The promoter P_{tac} has been shown to be a strong promoter driving the expression of reporter protein in *C. necator* under constitutive condition with 0.5% fructose as the carbon source [58]. Both P_{lac} and P_{phaC} are also able to drive the expression of reporter protein under similar

growth condition, however, the level of gene expression by these promoters were lower than that obtained with P_{tac}. A more recent study reported P_{j5} to be a stronger constitutive promoter than P_{tac} [59]. This information has guided rational engineering of promoters for gene expression in *C. necator* [57]. Specifically, incorporation of regulatory genetic elements (synthetic ribosomal binding site) and/or configuration alteration gave more than 100% change in relative promoter activity [57]. It is noteworthy that promoter activity is also dependent on plasmid backbone. For example, under pKRSF1010 backbone, P_{j5} had ~ 1.8-fold increase in activity compared to P_{g25} [59]. In contrast, under pBBR1MCS backbone, P_{g25} had ~ 1.5-fold increase in activity comparable to P_{j5} [57]. Although P_{tac}, P_{j5} and variant thereof were earlier established to be strong constitutive promoters for gene expression in *C. necator*, these promoters were compared with $P_{j23100PETRBS}$, extensively characterised in *E. coli* [60]. Gene expression driven by $P_{j23100PETRBS}$ gave higher normalised fluorescence across different reporters compared to that obtained with P_{tac} or P_{j5} promoters Chapter 4.

Isopropyl *β*-D-1-thiogalactopyranoside (IPTG) is not a good inducer of gene expression in *C. necator* [58]. This effect was attributed to the apparent lack of LacI homology in the bacterium [9], specifically the galactose permease gene *lacY* [51]. To overcome this, a *lacY* gene was codon-optimised for *C. necator* and the resulting promoter system, P_{lacUV5}, was inducible with IPTG [51]. However, the expression level was low comparable to that of other inducible systems, P_{BAD}, P_{BAD}T7 and P_{xylS}. Further, inducible expression system for *C necator* based on *lacI* and *cymR* were designed to be driven by P_{j5} promoter [61]. As expected, induction was achieved for the systems using their respective inducers, IPTG and *p*-cumate. Although induction with IPTG resulted in higher GFP expression, growth was significantly impaired under IPTG induction compared to *p*-cumate induction. Given that both systems were under the same promoter, P_{j5}, the impact on culture growth was ascribed to inducer mechanism of transport across cell membrane. IPTG is actively transported whilst *p*-cumate diffuses across cell membrane [61]. It appears P_{BAD} with L-arabinose as the inducer, is so far the best inducible system for *C. necator* application compared to other inducible systems [51,62]. The induction level by this promoter appears to be concentration dependent, with maximum induction at 0.1% L-arabinose [58]. More importantly, gene expression with P_{BAD} is tightly regulated in the absence of L-arabinose as the inducer.

Comparisons of fluorescence output of different promoters when reporter protein was the cargo, resulted in similar expression levels when the same promoters directed the expression of a biosynthetic gene [14,51,62]. Constitutive promoters tend to give higher fluorescence output than inducible promoters. Nevertheless, in most cases, this does not translate to higher productivity and growth rate when the goi encodes a biosynthetic pathway. For example, the

inducible promoter P_{BAD} allows for higher production of isopropanol and maximum growth rate compared to P_{lac} and P_{tac} [14]. It was observed that increasing constitutive promoter strength results in increase in productivity, which is accompanied with negative impacts on growth rate. This is not the same with the inducible promoter system, which results in better substrate consumption and conversion to final product. Under hydrocarbon synthesis, P_{BAD} system is a better promoter than P_{xyls} and P_{lacUV5} [51].

Table 1. 1 Range of promoters validated for use in *C. necator*.

Promoter	Origin relative to *C. necator* H16	Condition of activity	Inducer	Gene/reporter	Reference
P_{phaC1}, P_{rrsC}, P_{j5} and P_{g25}	Exogenous	Constitutive		*rfp*	[57]
TolC	Native	Inducible	Doxycycline	*gfp*	[63]
P_{araBAD}, P_{rhaBAD}, P_{cmt} and P_{acuRI}	Exogenous	Inducible	Arabinose, rhamnose, acrylate and cumate, respectively	*rfp*	[62]
P_{SH}	Native	Inducible	Hydrogenase de-repressing condition	*gfp*	[64]
P_{PhaC1}, P_{trc}, P_{lacUV5}, and P_{trp}	Exogenous			*Laz*	[65]
$P_{j5/lacI}$ and $P_{j5/cymR}$	Exogenous	Inducible	IPTG and *p*-cumate, respectively	*gfp*	[66]
P: *h22b*, *f30*, *de33*, *n25*, *n26*, *g25*, *k28a*, *T5*, *k28b*, *h207*, and *j5*	Exogenous	Constitutive		*gfp*	[59]
P_{lac}, P_{tac} / P_{BAD}	Exogenous	Constitutive/ Inducible	L-arabinose for P_{BAD}	Multiple genes for Isopropanol production	[14]
P_{rrsC} and P_{phaC1}	Native	Inducible	Anhydrotetracycline	*gfp*	[67]
P_{BAD}, P_{BAD}T7, $P_{xyls/PM}$, and P_{lacUV5}	Exogenous	Inducible	L-arabinose, m-toluic acid and IPTG, respectively	*rfp* and hydrocarbon genes (*aar* and *adc*)	[51]
P_{tac}, P_{lac}, and P_{BAD},	Exogenous	Constitutive except P_{BAD}		*gfp*	[58]
P_{phaC}, P_{phaP}	Native	Constitutive and inducible	L-arabinose for P_{BAD}	*gfp*	[58]

In summary, inducible promoters are preferred over constitutive promoters for controlled gene expression. However, it is important to consider: the effect of an inducer on growth rate, concentration needed for optimal induction, induction factor (maximum normalized fluorescence at highest inducer concentration per normalized fluorescence of uninduced sample) and whether the inducer will be co-utilised as a carbon source. There is a consensus that native promoters tend to exhibit less activity compared to non-native promoters [52,57,58,62].

1.6 Maintaining plasmid segregational stability in *C. necator*

One of the challenges facing propagation of foreign gene(s) carried on a plasmid vector in *C. necator* is plasmid loss due to low segregational stability. This occurs as a result of metabolic burden elicited by expressing a foreign gene(s), which in most cases significantly reduce cell growth leading to reduction in metabolic activity; this in turn will affect product yield. Plasmid stability is primarily a function of the *oriV* in relation to copy number, antibiotic marker and the goi. Some *oriV* have been reported to result in significant plasmid loss when used to direct heterologous protein expression in *C. necator* [68]. Although it is evident that addition of antibiotic in a cultivation medium will reduce plasmid loss, choice of such antibiotics is crucial. Kanamycin is so far the best antibiotic for maintaining plasmid stability in *C. necator*. The use of antibiotic such as chloramphenicol or tetracycline has been reported to increase plasmid loss [68,69,Chapter 4]. For example, the difference in plasmid loss when tetracycline was used as a selective pressure instead of kanamycin was 24% [69]. Also, 57.31 ± 17.19% difference in plasmid loss was observed with chloramphenicol as a resistant cassette instead of kanamycin [Chapter 4]. Interestingly, plasmid with complete pBBR1 replication machinery bearing kanamycin as the resistant marker was observed to retain segregational stability over several generations, in the absence of antibiotic pressure [Chapter 4].

Alternatively, any goi can be integrated in a host's genome/onto its chromosome to overcome the use of antibiotic as a selective pressure especially during large-scale fermentation. Moreover, the use of antibiotic to maintain the stability of an expression host during large-scale cultivation is often not economically competitive. However, this approach of chromosomal integration tends to result in low yield and productivity due to limited gene copy in comparison to plasmid-based expression system [68,69]. Having goi on plasmid, to replicate independently of *C. necator* replicons, exerts less metabolic burden on the bacterium and will allow for more gene copies, which will lead to improve product yield. Nevertheless, with a high copy replicon, the use of addiction system has proved effective in maintaining segregational stability under prolonged cultivation in the absence of antibiotic pressure [68,69].

Furthermore, RP4 partitioning system proved effective in maintaining plasmid stability in *C. necator* [59]. This partitioning system encodes post-segregational killing system (*parDE*)—a toxin (parE) and antitoxin (parD)—which favours the establishment of plasmid containing daughter cells and kills that of plasmid free cells by inhibiting protein synthesis [59,70]. In addition, RP4 operon (*parCBA*) encodes a nuclease (parB), a resolvase (parA) and a protein of unknown function (parC), which is suggested to aid in plasmid segregational stability to daughter cells [71,72]. It is noteworthy that plasmid copy number and host system determine the

stabilising effect of *parCBA/DE* [70]. Under low copy number condition, the stabilising effect of *parCBA* alone was considerably reduced, with *parDE* segregational killing having the major effect under this condition. Nonetheless, when the entire operon (*parCBA/DE*) is involved the stabilising effect was more effective notwithstanding plasmid copy number [70]. Application of RP4 partitioning system to *C. necator* resulted in < 100 stability for plasmid bearing pBRR1Rep and 100% for RP4, RSF1010, pSa and pSa variant [59]. Thus, it is likely that the lower stability obtained for plasmid with pBBR1Rep is perhaps due to *Rep* being a partial sequence of pBBR1 replication element or due to the negative interaction between plasmid with low copy number and partial partitioning system. Plasmid multimer resolution by the *parCBA* operon is not sufficient to stably maintain plasmid during long-term cultivation [70]. Therefore, a good choice of origin of replication, antibiotic resistance cassette and partitioning system are crucial to ensure stability of cloned gene in *C. necator*.

1.7 Transformation protocols

C. necator unlike *E. coli* is less competent to transformation by the widely known bacterial transformation methods: heat-shock, conjugation and electroporation. Transformation of *C. necator* is mostly achieved by means of conjugation (Fig 1.5A). This method of transformation relies on a donor bacterium, *E. coli* S17-1, to mate with *C. necator* in order to exchange genetic information. However, recent studies have shifted attention from conjugation to electroporation. Unlike conjugation, electroporation does not rely on a donor bacterium thus genetic information can be delivered directly to *C. necator*. Nonetheless, electroporation in *C. necator* is limited by plasmid backbone. pBBR1 plasmids (pBHR1 and pBBR1MCS-2) and their variants are the plasmids that are mostly used for *C. necator* studies. Electroporation success rate for pBHR1 is 100% whilst that for pBBR1MCS-2 is 20% (Fig. 1.5B). It is evident that the ineffective transformation of *C. necator* with pBBR1MCS-2 is due to the plasmid backbone, specifically due to the sequence of kanamycin resistant cassette Chapter 4.

Although electroporation success rate with pBHR1 is 100%, the transformation efficiency obtained for *C. necator* with this plasmid and other plasmids is low. Therefore, recent studies focused on improving electroporation transformation efficiency in *C. necator* [56,73]. Three major approaches have been employed to achieve this, namely optimisation of transformation protocols, host modification and plasmid engineering.

C. necator transformation via electroporation protocol was described in [74]. Pulse field strength (kV/cm), DNA concentration (µg), concentration of competent cells and cell concentration (optical density) at the time of harvest were investigated. Field strength of 11.5 kV/cm (1.15 kV or 2.3 kV for 1-mm or 2-mm cuvette, respectively) was reported as the optimum field strength, while harvesting cells at 0.8 OD_{600nm} yielded the best transformation efficiency. In addition, 1 *µ*g plasmid DNA was reported as the optimum DNA concentration. However, this is relative to the size of the plasmid used, pKT230. With pCAT plasmids, ~ 50 ng of plasmid DNA is enough to achieve higher transformation efficiency Chapter 4.

Transformation buffer containing 0.2 M sucrose resulted in 1000-fold increase in transformation efficiency in comparison to buffer with 10% glycerol, whilst chemical treatment with 50 mM $CaCl_2$ for 15 min improved efficiency by approximately two-fold [73]. Harvesting cells at mid exponential growth phase (0.4–0.8 OD_{600nm}) improved efficiency [73,74]. Higher concentration of cells gives better transformation efficiency compared to lower concentration, as there are more cells to transform [73]. Thus, harvesting cells at late exponential to early stationary phase (beyond 0.8 OD_{600nm}) will likely not have significant negative impact on transformation efficiency compared to harvesting cells during early exponential growth phase, when cells are fewer.

Modification of the *C. necator* genome is gaining attention as a method of improving gene delivery to the host. Deletion of restriction endonuclease genes (*H16_A0006* and *H16_A0008-9*) significantly improved transformation efficiency, by ~ 1697-fold, compared to the wild type harbouring such endonucleases [56]. Unlike host modification, plasmid engineering does not require disruption of any host gene in order to obtain high transformation efficiency. This offers advantage of working with the wild type strain. Re-engineering of broad host range (pBBR1) plasmids backbone resulted in significant increase, more than 3000-fold, in transformation efficiency compared to the efficiency obtained with existing plasmids: pBHR1, pBBR1MCS-2 and their variants Chapter 4. Interestingly, *C. necator* can be transformed by heat-shock method; this adds to the toolbox for the bacterium Chapter 3.

There is room for further improvement in transformation efficiency of *C. necator*. For a model microbial chassis, *E. coli*, preparation of electrocompetent cells at room temperature (24–28 ºC) considerably improved efficiency compared to ice-cold competent cells [75]. Room temperature cells tend to be smoother and less shrunken in comparison to cells maintained in ice-cold condition. The latter condition has been routinely used for *C. necator*. Room temperature transformation can be replicated in *C. necator* to observe the effect on transformation efficiency. Revisiting transformation protocols with high transforming

plasmids, such as pCAT plasmids, may result in further increase in transformation efficiency of *C. necator* thus making the process of transforming the bacterium more efficient.

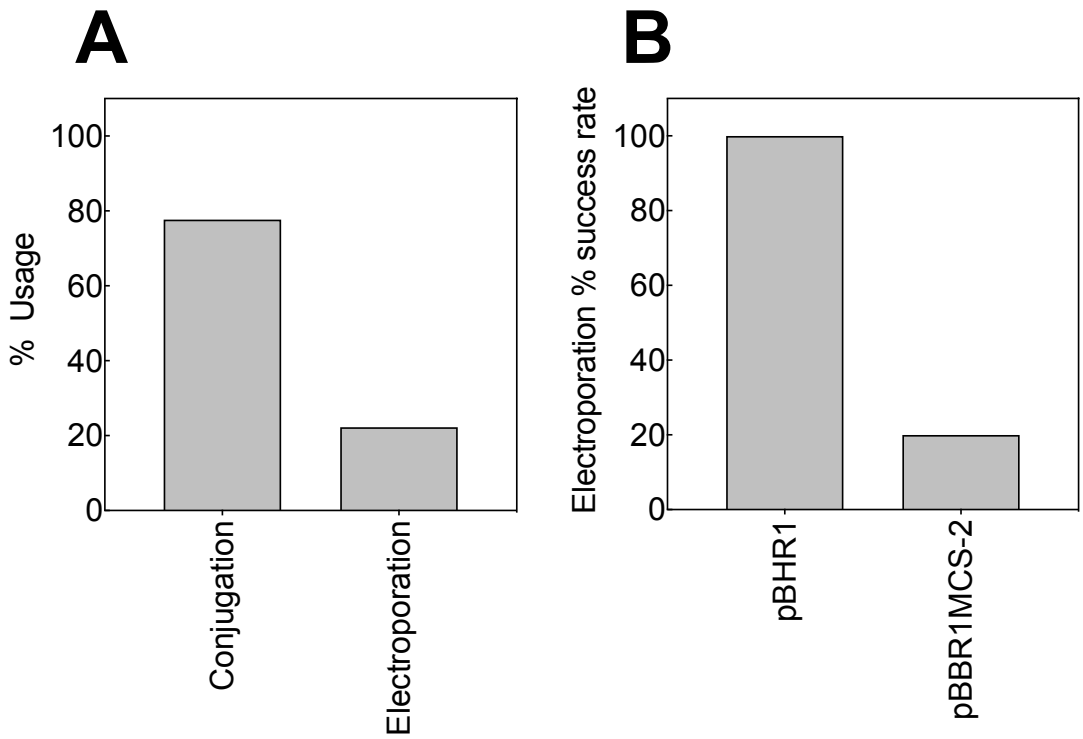


Fig. 1. 5 Method of transforming *C. necator*.

A. Percentage transformation of *C. necator* by conjugation or electroporation. **B**. Electroporation success rate with the two most commonly used broad host range plasmids for *C. necator*. Data are from more than 80 genetic related studies involving *C. necator* H16 as the host from 1981 to 2019.

1.8 Targeted gene replacement in *C. necator*

Gene deletion in *C. necator* is achieved by homologous recombination. This method of recombination relies on the use of a suicide vector carrying essential bioparts: narrow host range *ori*, *sacB*, antibiotic resistant marker and goi—flanked by DNA sequences homologous to that of the host sequences. Mutants with replaced genes (homogenotes) are obtained following two successive recombination events, whereas genome integrated mutants (heterogenotes) are obtained from single recombination event. Levansucrase gene (*sacB*) from *Bacillus subtilis* is used as a counter selection marker for *C. necator* and other Gram-negative bacteria homogenotes [76–78]. In the first recombination event, mutants are selected based on resistance to antibiotic on nutrient rich agar plates [79]. Subsequently, homogenotes are selected in a second recombination event based on resistance to sucrose in an antibiotic- and NaCl-free medium [77]. Although homogenotes can be selected in a single

recombination event, two-step recombination event increases the chance of obtaining double crossover mutants, homogenotes [76]. This approach has successfully been used to achieve gene replacement in *C. necator* [11,12,16,44,65,80]. Nevertheless, false homogenotes may arise due to mutation in *sacB* [76]. Moreover, conjugation is predominantly used to deliver suicide plasmid to *C. necator*, which makes it more time consuming. Given that suicide plasmids with flanking homologous regions to target gene in *C. necator* are constructed with narrow host range origin of replication [76,77,81], the resulting plasmids after amplification in *E. coli* can be delivered directly to *C. necator* via electroporation. This will increase the efficiency of obtaining *C. necator* recombinants. Additionally, it will reduce any chance of carrying *E. coli* S17-1 through to subsequent recombination events, as plasmids with narrow host range origin are unable to direct autonomous replication in *C. necator*. Hypothetically, suicide plasmid for gene replacement in *C. necator* can be assembled without narrow host range origin of replication and delivered directly to the bacterium thus increasing the efficiency of the technique.

A markerless gene deletion system for *C. nectaor* was developed based on group II intron [82]. This technique depends minimally on the host and uses a DNA integration mechanism, retrohoming, mediated by a ribonucleoprotein formed during RNA splicing. However, the reliance of the technique on computer algorithm to calculate optimal site for the integration based on probability makes it less efficient. The knockout efficiency obtained with this technique is 12.5%.

High efficient electroporation based genome editing via clustered regularly interspaced short palindromic repeat (CRISPR-Cas9) system has been demonstrated in *C. necator* [56]. Nonetheless, this was achieved using a different plasmid backbone (pBBR1MCS-2), and the resulting transformation efficiencies were lower compared with that obtained for the plasmid counterpart, without Cas9 or Cpf1 genes. The efficiency of this system can be improved by replacing pBBR1MCS-2 kanamycin resistant sequence with that from pBHR1.

1.9 Outlook

Cupriavidus necator holds great promise as a microbial chassis for harnessing renewable resources for biosynthesis of high-valued compounds. There is a steady rise in studies involving *C. necator* (Fig. 1.6A). The majority of the studies are on biopolymers and enzymes, with fewer studies on platform chemicals (Fig. 1.6B). However, studies on tool development and platform chemical have the highest turnover in the last two decades (Fig. 1.6C). Bioplastic accumulation in the bacterium is receiving significant attention with few

studies synthesizing hetero-monomers other than homo-monomers [20,21,65,83]. It will be interesting to advance bioplastic production in *C. necator* to producing aromatic hetero- and/or homo-monomers, using cheap and abundant raw materials. This will improve the material properties of bioplastic from *C. necator* and will reduce the cost of raw materials, especially substrate needed for bioplastic production in the bacterium. Such monomers will be eco-friendly starting materials for the synthesis of polyester thermoplastic polymers, which currently are derived from petrochemicals.

One of the interesting features of *C. necator* H16 that is yet to be explored is, its bioremediation potentials. The bacterium is well equipped to grow on range of aromatics [9], and can grow on benzoates, phenols, *p*-cresol and other similar aromatic compounds [84]. However its ability to be deployed as a bioremediator is shadowed by other *C. necator* species, specifically strain JMP134 [35,85]. More studies on *C. necator* H16 utilisation of aromatics will provide insight into genes that play significant role during the process of aromatic compound degradation in the bacterium. Further, it will help elucidate pathways, potentially new ones, for catabolism of aromatic compound in the bacterium thus making it more appealing not only for biosynthesis but also for bioremediation.

Tool development for precise genome editing using CRISPR system will improve the efficiency of genome engineering in *C. necator*. Potentially, multiplex genome engineering can be achieved in a single transformation event rather than in a time-consuming sequential transformation. This will significantly accelerate metabolic engineering and synthetic biology studies in *C. necator* for production of yet to conceive high-valued compound. As with other microbial chassis, tolerance to biofuels is a bottleneck and tends to affect yield during biosynthesis. With the possibility of transforming libraries directly to *C. necator* [Chapter 4], directed evolutionary approach based on error-prone PCR or custom-designed DNA oligo library pool [86] can be employed to generate *C. necator* mutant with improved biosynthetic features. Aware that the industrial features of *C. necator* are duplicated and distributed across its replicons, minimal genome of the bacterium based on its unique features can be designed and constructed. The resulting genome can be installed in another receptive host [87]. Such approach will potentially reduce the metabolic burden on the bacterium contributed by its complex genome.

This study provides a comprehensive view on the existing genetic toolkit for *C. necator*. The information provided here would be crucial for accelerating the development and deployment of *C. necator* as a microbial chassis for biosynthesis and bioremediation.

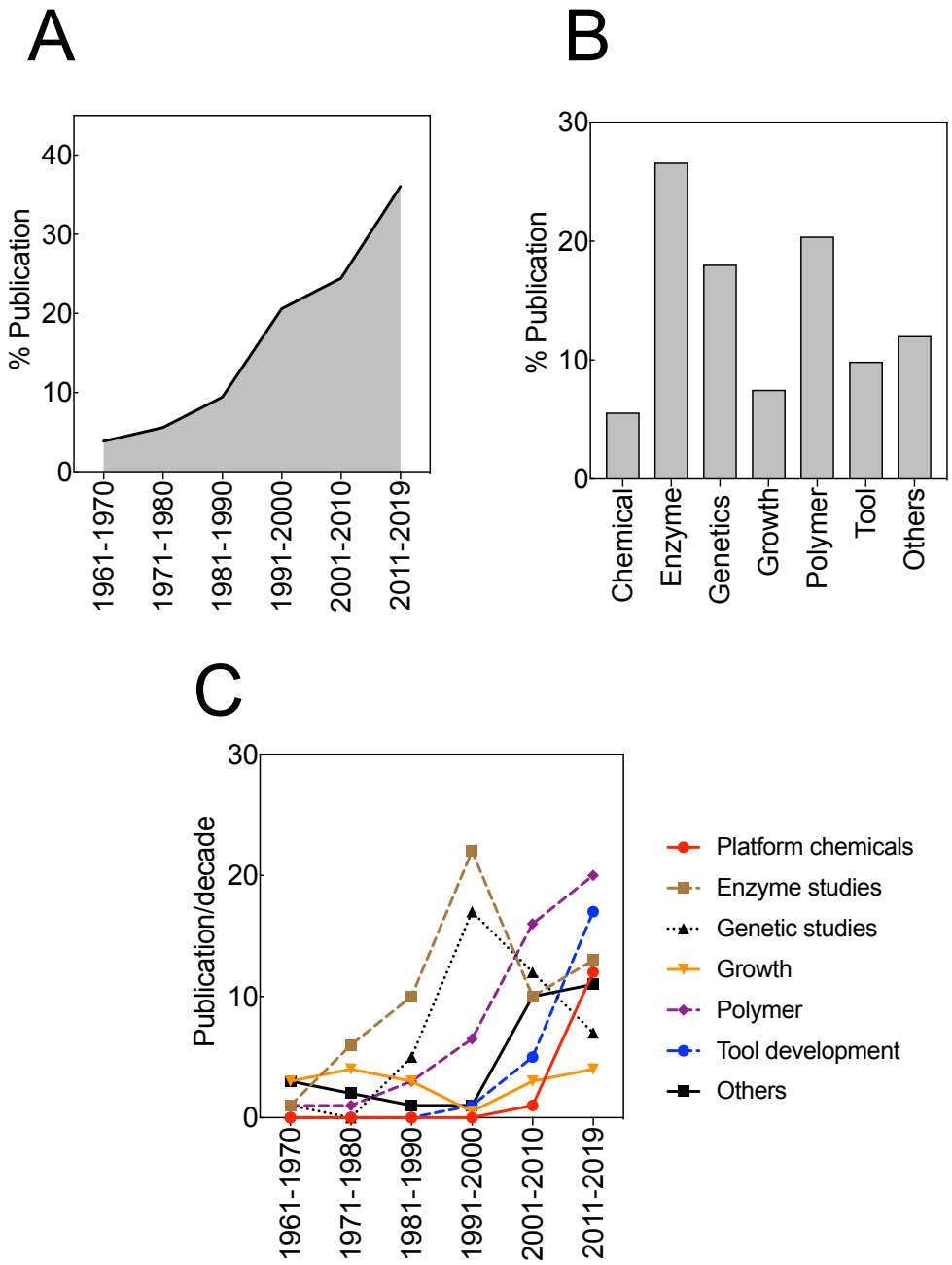


Fig. 1. 6 Progress on *C. necator* studies since discovery.

A. General percentage publication over decades. **B**. Cumulative percentage publication with respect to research interest. **C**. Comparison of research output from different *C. necator* research interest. Data are from more than 230 research article on *C. necator* from 1961 to 2019.

1.10 Rationale and aim of the study

With the understanding that bioengineering applications in *C. necator* are hindered by lack of genetic toolkit, this study is designed to characterise some of the existing toolkit to make them more suitable for *C. necator*. The aim of this study, therefore, is to develop and expand the genetic toolkit for *C. necator* to facilitate rapid and controlled bioengineering applications in the bacterium, for predictable output.

Objectives

In line with the aim, the objectives of the study are:

- Understand how *C. necator* responds to components of defined media.
- Build a growth model predicting *C. necator* growth in defined media.
- Characterise existing plasmids biological parts (bioparts) for improved transformation efficiency in *C. necator.*
- Modularise and minimise plasmid parts for efficient and high throughput cloning in *C. necator*.
- Make *C. necator* a direct recipient of genetic materials, without passing through *E. coli* as an intermediary host.
- Demonstrate application of the toolkit by engineering *C. necator* for improved tolerance to biofuels.

Hypotheses

In this study, the following general hypotheses were tested:

- *C. necator* tends to have extended lag phase of growth in a defined medium. Addition of few amino acids to *C. necator* defined growth medium will improve growth rate and shortens lag phase of growth.
- *C. necator* is susceptible to transformation with limited number of existing plasmids. Moreover, the transformation efficiency obtained is low. Incompatibility of bioparts is responsible for the low transformation efficiency observed for *C. necator* with existing plasmids.
- Reengineering of existing plasmids will improve transformation efficiency in *C. necator*.
- Mutation in *C. necator* σ^{70} (*rpoD*) will improve tolerance to chemicals.

1.11 Thesis structure

Chapter one of this study is a literature review highlighting progress on the development of genetic toolkit for *C. necator*. It covered the nomenclature of *C. necator* and distribution of key metabolic functions across replicons (chromosomes and plasmid) in relation to its applications. Limitations of transformation protocols with attention on bioparts incompatibility were discussed. It also highlighted the promoters suitable for application in the bacterium.

Chapter two focuses on building a predictive growth model for *C. necator*. Growth medium is an essential tool for any microbial chassis. Nevertheless, for a bioplastic producing bacterium such as *C. necator*, little is known on how the bacterium responds to different components in defined media. Understanding how components of defined media affect *C. necator* growth will contribute immensely to scaling biomanufacturing application with the bacterium. A manuscript from this chapter is publicly available on bioRxiv (https://doi.org/10.1101/548891), and it is also provided in the appendix.

Chapter three focuses on characterising existing plasmid bioparts for efficient transformation of *C. necator*. Another essential tool for genetic manipulation of a microbial host is plasmid. This tool is poorly characterised for *C. necator* thus making the first step of bioengineering work (cloning) in the bacterium a daunting task. Therefore, it is crucial to characterise bioparts for *C. necator* to make the bacterium more appealing for bioengineering applications. This chapter set the scene for chapter four, which focuses on building modular minimal plasmids for the bacterium. The parallel assembly protocol demonstrated in this chapter—with individual biopart characterised in *C. necator*—will unlock the major limitation associated with cloning in the bacterium.

Lastly, chapter five demonstrates the application of the toolkit established by engineering *C. necator* for improved tolerance to chemicals using global transcription machinery engineering (gTME) approach. Microbes are currently sort after as production platform for solvents including biofuels. However, optimising the titre, yield and productivity is limited by the tolerance level of a given microbial chassis. Hence, the use of a chassis organism tolerant to the compound of interest would improve productivity. Such chassis organism can be expeditiously engineered by delivering mutated library resulting from amplification and high-throughput assembly directly to the host. This is followed by screening the resulting transformants for desired phenotype.

The findings of this study will guide the development of more tools for *C. necator* and other non-model bacterial chassis. More importantly, it would expedite metabolic engineering and synthetic biology applications in *C. necator*.

Chapter 2.

Statistical Design of Experiments (DoE) Reveal Medium Components Affecting the Growth of *Cupriavidus necator* H16

2.1 Introduction

The flexible bioenergetics features of *Cupriavidus necator* make the bacterium a promising microbial chassis to produce commodity chemicals and fuels using renewable resources. One of the prominent features of *C. necator*, biopolymer accumulation, occurs under growth limiting condition—notably nitrogen, phosphorous or oxygen limiting condition(s). When the limiting condition is restored, the accumulated biopolymer pool is depolymerised to simple carbon, to be used as a source of carbon for growth [20,42,50,88]. Industrial applications of *C. necator* have been demonstrated in the biosynthesis of high valued products ranging from alcohols [11–13,42], fatty acids [16–18], alkanes [19], and enzymes [22]. The production of branched chain alcohols from *C. necator* by electrosynthesis (i.e. electrical powered cellular synthesis) demonstrates the value of the bacterium as a chassis for producing valued product using renewable resources [11].

Underpinning the development of *C. necator* as an industrial chassis is the requirement for chemically defined media (CDM). CDM are important tools to enable experimental reproducibility, to reliably characterise genetic parts and devices, to determine genotype by environment interactions and to facilitate fundamental research of its microbial physiology that underpin bioengineering efforts. Unlike complex growth medium where the composition of some components are not completely known [89], CDM components and compositions are known exactly thus reducing variability during large-scale cultivation. This makes it easy to quantify the impact of components of CDM on growth and productivity. The use of a complex growth medium during production process impacts negatively on productivity due to the potential release of some microbial substrates from pool of polymers and precursors, thus rendering the computation of substrate turnover and productivity difficult [90]. While different chemically defined media have been deployed for the cultivation of *C. necator* [12,13,16,20,43,91,92], there is no consensus regarding the components that are required, the concentration of each component, or how component interacts to impact on growth (Table 2.1).

Statistical Design of Experiments (DoE) is crucial for modelling a system response. It allows the systematic examination of multiple variables simultaneously to gain maximum information

on a system by relying on fewer experimental runs unlike traditional one factor at a time (OFAT) approach [93]. Effective DoE approaches can result in a model that can predict the outcome of a system at different scenarios [94]. Importantly, with DoE, interactions between factors can be revealed, which in most cases are key determinant of a system output [95]. Despite its origins within the biological sciences [96], DoE is not commonly applied in biological processes. However, it has recently found value in the optimisation of metabolic pathways [97,98], cell-free protein synthesis reactions [99] and codon-usage algorithms [100], the modelling of genotype-by-genotype and genotype-by-environment interactions in yeast metabolism [101], optimisation of fusion protein in *Pichia pastoris* [102] and in enzyme repurposing [103]. The lack of a consensus in *C. necato*r growth medium makes it difficult understanding which components play important role in defined media formulations for this industrially relevant bacterium. Therefore, this study aims at building a robust predictive growth model for *C. necator* using a chemically defined medium. The ability to predict the outcome of microbial growth under different growth scenarios is vital in industrial fermentation processes and other biotechnological applications. Understanding the impact of each component of a CDM on *C. necator* growth is a fundamental tool for the controlled exploration of the biotechnological potentials of the bacterium.

Table 2. 1 Previously described chemically defined media for *C. necator*.

All concentrations are g/L unless otherwise indicated. Starting medium for DoE investigation ([a]).

Carbon	Phosphate buffer	Potassium	Magnesium	Calcium	Nitrogen	Trace elements	Reference
50 Fructose	1.5 Na_2HPO_4. $12H_2O$, 0.25 KH_2PO_4,		0.75 $MgSO_4.7H_2O$	0.015 $CaCl_2$	2.39 $(NH_4)_2SO_4$, 0.285 nitrilotriacetic acid, 28% 0.9 $Fe(NH_4)_2$(citrate)	0.00045 H_3BO_3, 0.0003 $CoCl_2$, 0.00015 $ZnSO_4.7H_2O$, 0.000045 $MnCl_2.4H_2O$, 0.000045 NaMoO4.$2H_2O$, 0.00003 $NiCl_2.6H_2O$ and 0.000015$CuSO_4$. $5H_2O$	42
0.5-1% Fructose	12 ml 1.1 M H3PO_4	0.45			0.1 or 0.01% NH_4Cl	24 mL/trace: 2.6 $CaCl_2.2H_2O$ 0.1 $MnSO_4.H_2O$, 0.1 ZnSO4.$7H_2O$, 0.02 $CuSO_4.5H_2O$, 0.015 $FeSO_4.7H_2O$	16
10-20 Fructose	1 KH_2PO_4; 11.1 Na_2HPO_4. $12H_2O$		0.2 $MgSO_4$		3 $(NH_4)_2SO_4$	9.7 $FeCl_3$, 7.8 $CaCl_2$, 0.156 $CuSO_4.5H_2O$), 0.119 $CoCl_2$, 0.118 $NiCl_2$ and 0.062 $CrCl_2$	43
2% v/v biodiesel-driven glycerol, 2 % w/v glycerol bottom and 1% v/v free fatty acids	6.7 Na_2HPO_4. $2H_2O$, 1.5 KH_2PO_4,		0.2 $MgSO_4.7H_2O$	0.01 $CaCl_2$.$2H_2O$	1 $(NH_4)_2SO_4$, 0.06 $Fe(NH_4)2$(citrate)	0.3 H_3BO_3, 0.2 $CoCl_2$, 0.1 $ZnSO_4.7H_2O$, 0.03 $MnCl_2.4H_2O$, 0.02 NaMoO4.$2H_2O$, 0.02 $NiCl_2.6H_2O$ and 0.01$CuSO_4.5H_2O$	92
20 Fructose	1.51 Na_2HPO_4, 2.65 KH_2PO_4,		0.3 $MgSO_4.7H_2O$			1.97 $FeCl_3$, 9.7 $CaCl_2.2H_2O$, 0.156 $CuSO_4.5H_2O$, 0.184 $CoCl_2.6H_2O$, 0.118 $NiCl_2.6H_2O$ and 0.062 $CrCl_3$	20
[a]2% Fructose, 2% sodium gluconate, 1% palm oil or 0.5% Tween-60	4 NaH_2PO_4, 4.6 Na_2HPO_4	0.45 K_2SO_4	0.39 $MgSO_4$	0.062 $CaCl_2$	0.5 NH_4Cl	15 $FeSO_4.7H_2O$, 2.4 $MnSO_4.H_2O$, 2.4 ZnSO4.$7H_2O$, 0.48 $CuSO_4.5H_2O$	12,44
2-4 Fructose, 2 ml glycerol	1.5 KH_2PO_4, 9 Na_2HPO_4		0.2 $MgSO_4.7H_2O$	0.01 $CaCl_2$.$2H_2O$	2 NH_4Cl	0.05 $FeCl_3.6H_2O$, 0.019-0.190 $NiCl_2.6H_2O$	13

2.2 Materials and Methods

2.2.1 Bacterial strain

Cupriavidus necator H16 (DSM 428) was obtained from Deutsche Sammlung von Mikrooganismen und Zellkulturen GmbH, German Collection of Microorganisms and Cell Cultures (DSMZ). The bacterial strain was resuscitated on nutrient agar (peptone 5 g/L, and meat extract 3 g/L) according to supplier's instruction and incubated at 30°C for 48 h.

2.2.2 Media components

Carbon sources (glucose, fructose, glycerol and sucrose), salts (except $MgSO_4.H_2O$ and NH_4Cl), trace metals, amino acids (histidine, leucine and arginine) and some vitamins (thiamine, niacin, pantothenic acid) were obtained from Sigma-Aldrich. The remainder of the vitamins, $MgSO_4.H_2O$, NH_4Cl and methionine were obtained from Duchefa Biochemie B.V., BDH chemicals and Formedium, respectively.

2.2.3 Media preparation

A comparison of the literature for the use of defined media for *C. necator* growth identified variety in both the nature and concentrations of macroelements and trace elements required for robust cell growth (Table 2.1). Therefore, components that served as factors for the DoE were carefully chosen based on knowledge on general nutritional requirement for bacteria. The macroelements considered were C, N, O, P, Ca, Mg, S, whilst trace elements were Cu, Zn, Fe and Mn. These basic components correspond to the components of one of the existing CDM for *C. necator* [44]. This medium (in g/L: 20 fructose, 4 NaH_2PO_4, 4.6 Na_2HPO_4, 0.45 K_2SO_4, 0.39 $MgSO_4$, 0.062 $CaCl_2$, 0.5 NH_4Cl and 1 mL of trace solution containing in g/L: 15 $FeSO_4.7H_2O$, 2.4 $MnSO_4.H_2O$, 2.4 $ZnSO_4.7H_2O$ and 0.48 $CuSO_4.5H_2O$) served as a starting point for the investigation. Additionally, few amino acids and vitamins considered important for microbial growth [104,105] were added as factors to determine their effect on *C. necator* growth. Because adjusting the pH of media might result in the addition of extra ions, which in turn might impact on interpretation of how components contribute to growth, pH of each medium was not adjusted but were measured when appropriate. *C. necator* has broad range of applications, ranging from enzymes, platform chemicals and biopolymer syntheses; however, optical density (OD_{600nm}) was chosen as the output to gain detailed understanding on how the bacterium responds to different nutritional conditions. Such understanding would guide other process improvement involving *C. necator* as a production host.

Solutions of glucose, fructose, sucrose, vitamins, amino acid and each trace metal solution were filter sterilised through 0.22 μm filter. Glycerol, NaH_2PO_4, Na_2HPO_4, $MgSO_4.H_2O$,

NH_4Cl, K_2SO_4, and $CaCl_2.2H_2O$ solutions were autoclaved. The trace solution was made of $CuSO_4.5H_2O$ (dissolved in 0.1M HCl), $FeSO_4.7H_2O$ (freshly prepared during each trace reconstitution from individual stock), $MnSO_4.H_2O$, and $ZnSO_4.7H_2O$. Amino acid stock solution contained: arginine, histidine, leucine, and methionine, while vitamin stock solution contained: folic acid, niacin, nicotinamide, pantothenic acid, pyridoxine, riboflavin and thiamine. Subsequently, medium components were added from individual stock solutions except for trace elements, which were added from reconstituted working solution. Forty-eight well plates were used in all small-scale trials, and medium reconstitution in each well was carried out using an automated liquid handling system (Eppendorf epMotion M5073). All stock solutions were prepared using water as a solvent and were further diluted in sterile distilled water.

2.2.4 Inoculum preparation and growth measurement

For each experiment, 48 h colonies from LB agar were washed twice in sterile distilled water and diluted to working inoculum concentration, ~ 10^8 cfu/mL. The inoculum was further diluted 1:100 in wells containing medium. Plates were incubated in a shaking incubator at 30°C, 170 rpm. Optical density (OD) at 600 nm was measured every 24 h using a Varioskan LUX™ Multimode Microplate reader (Thermo Scientific).

2.2.5 Batch cultivation in a bioreactor system

Large-scale cultivations were carried out in a batch mode with 1 L chemically defined media contained in 2 L capacity fermentors (Applikon ADI fermentation system). During the cultivations, pH was maintained above 4.5, temperature at 30ºC, agitation at 200 rpm. Dissolved oxygen (dO_2) was maintained above 20% [14,42] with airflow rate at 1 vvm (volume of air per volume of medium). No anti-foam or base was added throughout the cultivation. Media for starter cultures were of the same formulations with that used in fermentors and were prepared by growing 100 mL culture (in 250 mL non-baffled flasks) to late exponential growth phase, 48 h at 30°C, 200 rpm. Following polarisation and calibration of dO_2 probe, fermentors were inoculated with 10 mL (48 h) starter cultures and cultivations monitored online and offline over 72 h. Samples were taken every 24 h for offline analysis, growth measurement (OD_{600nm}).

2.2.6 Data analyses

Statistical Design of Experiments (DoE) were created using JMP Pro statistical software (version 13.0) and data from each experiment was analysed using the same software. Graphics were generated using GraphPad Prism 7.0.

2.3 Results

2.3.1 Scoping of ingredients in chemically defined media

The preliminary phase of the experiment was to determine whether OD_{600nm} is appropriate for determining *C. necator* growth. To establish this, the bacterium was cultivated in a rich medium (LB broth), and correlation between the colony-forming units (cfu/mL) and OD_{600nm} determined (Fig 2.1). During exponential growth in LB broth, *C. necator* cfu/mL (viable cells) was found to correlate well with its optical density (OD_{600nm}).

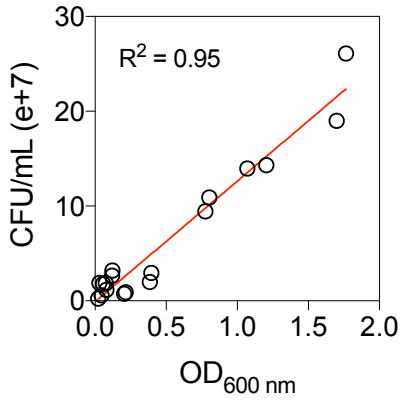


Fig. 2. 1 Optical density as surrogate for determining *C. necator* growth characteristics.

Colony-forming units (CFU) per mL against OD_{600nm}. Data are for *n* = 3 biological replicates.

Next, one of the commonly used CDM for *C. necator* [44] (Table 2.1) was chosen as a starting point for the investigation. An initial scoping trial was carried out using fructose, glucose, glycerol or sucrose to identify a principle carbon source to be used for subsequent experiment and to determine the range of concentrations to be tested for other components. Four scoping trials were performed, one each at low and high concentrations of media components, and two trials at the midpoint values between the two extremes (Table 2.2). Cultures were grown in 1 mL volumes in a 48-well plate for 120 h, at 30°C, 170 rpm.

Table 2. 2 Scoping trial design.

Media	Carbon	NaH_2PO_4	Na_2HPO_4	K_2SO_4	$MgSO_4$	$CaCl_2$	NH_4Cl	T.E.	A.A.	Vit.
Low	0.5	0.1	0.1	0.01	0.01	0.01	0.01	0.01	0.05	0.01
Medium	20.25	3.05	3.05	2.41	1.41	0.41	2.41	2.41	9.53	2.41
High	40	6	6	4.8	2.8	0.8	4.8	4.8	19	4.8

A scoping trial was carried out to gain understanding on the impact of range of media compositions on *C. necator* growth prior to embarking on formal design of experiments. All concentrations are in g/L except trace elements, amino acids and vitamins, which are mL/L. Trace element working concentration contained (g/L): 15 $FeSO_4.7H_2O$, 2.4 $MnSO_4.H_2O$, 2.4 $ZnSO_4.7H_2O$, and 0.48 $CuSO_4.5H_2O$. A 100x stock amino acid mix contained (g/L): 12.9 arginine, and 10 each of histidine, leucine and methionine. A 1000x vitamin stock contained (g/L): 0.1 pyridoxine, 0.02 folic acid, 0.05 each of thiamine, riboflavin, niacin, pantothenic acid and nicotinamide. Abbreviations: T.E. trace element mixture; A.A., amino acid mixture; Vit., vitamin mixture. Carbon: glucose, glycerol, fructose or sucrose. The medium trial was performed in duplicate.

At the range of concentration tested for all components, except the carbon sources, fructose supported the best growth, whilst glucose, glycerol and sucrose supported little or no growth at all concentrations (Fig. 2.2A-D). The OD_{600nm} for the two midpoint experiments showed a peak between 72–96 h and plateaued afterward. From these scoping trails, it was established that fructose would be the principle carbon source for the investigation, OD_{600nm} was an appropriate measure of cell growth, growth assays in 1 mL volumes in a 48-well plate were appropriate for subsequent experiments and recording OD_{600nm} at 72 h provides a good balance between measuring growth rate and peak culture density. These were important to establish the parallel experimental methods that underpin the approach.

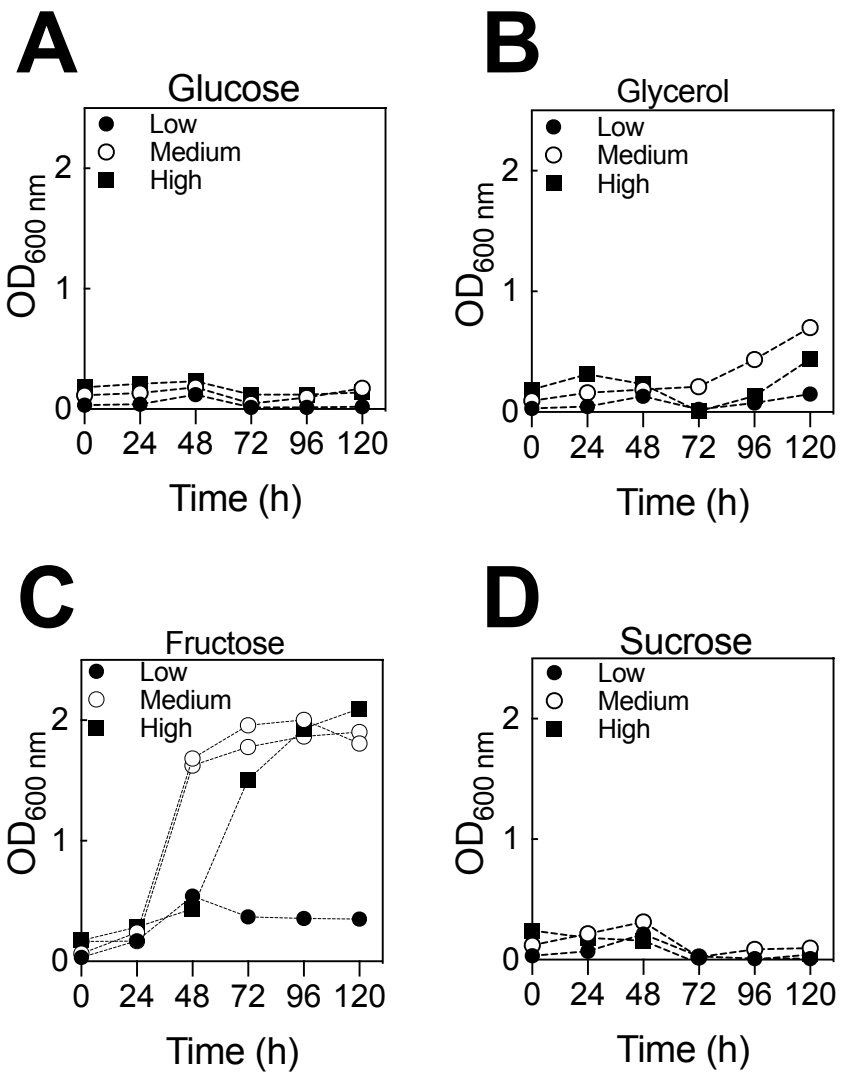


Fig. 2. 2 Growth of *C. necator* in 48-well plate format with different carbon sources.

A. Glucose. **B.** Glycerol. **C.** Fructose. **D.** Sucrose. Each experiment was carried out at low, medium (n = 2) or high concentrations of each media component (details can be found in Table 2.2).

2.3.2 Identifying main ingredients in chemically defined media

Having established key parameters for the investigation, an initial definitive screening design (DSD1) array was built based on 10 media components to identify the main effects of each component on growth (Table 2.3). DSDs are highly efficient experimental designs in which all main effects can be estimated independently of other main effects and all possible two-way interactions. The requisite variant media compositions were assembled in a 48-well plate using a liquid-handling robot. Cell growth in these variant media was monitored as previously described. The resulting data were analysed using both the Definitive Screening and the Two-Level Screening platforms in JMP Pro 13.0. Both analyses indicated that for high growth

to be achieved, fructose, $CaCl_2$, and amino acids must be maintained at high concentration whilst Na_2HPO_4 and trace elements must be maintained at low concentration (Fig. 2.3A). These effects were deemed statistically significant (t < -1.65 or > 1.65). Factors such as NaH_2PO_4, K_2SO_4, $MgSO_4$, NH_4Cl and vitamins were not found to have statistically significant effects under the concentrations tested.

In addition, these analyses (Definitive Screening and the Two-Level Screening) also highlighted several two-way interactions between components that are critical to growth. Factor interactions with significant positive impacts on growth at high concentration were between amino acids and fructose; fructose and $CaCl_2$; amino acid and trace elements, whilst an interaction between Na_2HPO_4 and fructose was identified as being detrimental to growth at high concentration (Fig. 2.3B). Two non-linear responses (fructose by fructose and $CaCl_2$ by $CaCl_2$) were also identified as having a negative impact on growth at 72 h period. However, while DSDs are efficient arrays for de-aliasing main effects from other main effects and from second-order interactions, second-order interactions remain partially aliased with themselves. Thus, attributing effects to specific second-order interactions was not attempted at this stage.

Beside the main effect of each component (factor), and interactions between some components of media, a closer look at the mean growth distribution between the settings or levels (low, mid and high) showed interesting responses (Fig. 2.3C). For example, the spread of growth for fructose showed a non-linear (quadratic) effect, indicating that the optimum fructose concentration that will result in robust and less variable growth lies above 0.5 g/L and below 40 g/L. Disodium phosphate (Na_2HPO_4) showed a slightly non-linear effect, with growth at low setting reaching higher OD_{600nm} than at high setting. Similarly, $CaCl_2$ showed a non-linear effect; the optimum concentration range for robust growth appears to lie above the low setting (0.01 g/L) and below the high setting (0.8 g/L). Trace element displayed a weak non-linear effect. Addition of few amino acids showed a positive linear effect, indicating that the optimum concentration range for amino acids is yet to be attained. Nevertheless, this was not attempted given that addition of more amino acids will contribute to extra carbon and nitrogen sources. The non-significant factors displayed linear effect, except $MgSO_4$. Therefore, this initial DSD needs augmentation to establish an optimum concentration range for each component, especially components having the most significant effect on growth.

Table 2. 3 Definitive screening design array.

Media	Fru.	NaH_2PO_4	Na_2HPO_4	K_2SO_4	$MgSO_4$	$CaCl_2$	NH_4Cl	T.E.	A.A.	Vit.	OD1	OD2
1	20.25	0.1	0.1	0.01	0.01	0.01	0.01	0.01	0.05	0.01	1.06	1.37
2	20.25	6	6	4.8	2.8	0.8	4.8	4.8	19	4.8	1.14	1.62
3	0.5	3.05	0.1	4.8	0.01	0.01	0.01	4.8	19	4.8	0.52	0.47
4	40	3.05	6	0.01	2.8	0.8	4.8	0.01	0.05	0.01	0.09	0.78
5	0.5	6	3.05	0.01	2.8	0.01	0.01	0.01	19	4.8	0.77	0.71
6	40	0.1	3.05	4.8	0.01	0.8	4.8	4.8	0.05	0.01	-	-
7	40	6	0.1	2.41	2.8	0.01	4.8	4.8	19	0.01	0.93	0.69
8	0.5	0.1	6	2.41	0.01	0.8	0.01	0.01	0.05	4.8	0.11	0.28
9	0.5	6	0.1	4.8	1.41	0.01	4.8	0.01	0.05	0.01	0.80	0.64
10	40	0.1	6	0.01	1.41	0.8	0.01	4.8	19	4.8	1.62	1.48
11	0.5	6	6	0.01	2.8	0.41	0.01	4.8	0.05	0.01	0.29	0.32
12	40	0.1	0.1	4.8	0.01	0.41	4.8	0.01	19	4.8	1.86	1.41
13	0.5	6	6	4.8	0.01	0.8	2.41	0.01	19	0.01	0.41	0.55
14	40	0.1	0.1	0.01	2.8	0.01	2.41	4.8	0.05	4.8	-	0.01
15	0.5	0.1	6	4.8	2.8	0.01	4.8	2.41	0.05	4.8	0.17	0.43
16	40	6	0.1	0.01	0.01	0.8	0.01	2.41	19	0.01	1.69	1.67
17	0.5	0.1	0.1	4.8	2.8	0.8	0.01	4.8	9.53	0.01	0.17	0.28
18	40	6	6	0.01	0.01	0.01	4.8	0.01	9.53	4.8	-	0.19
19	0.5	0.1	0.1	0.01	2.8	0.8	4.8	0.01	19	2.41	0.48	0.70
20	40	6	6	4.8	0.01	0.01	0.01	4.8	0.05	2.41	-	0.04
21	20.25	3.05	3.05	2.41	1.41	0.41	2.41	2.41	9.53	2.41	1.76	1.70
22	0.5	0.1	6	0.01	0.01	0.01	4.8	4.8	19	0.01	0.86	0.71
23	0.5	6	0.1	0.01	0.01	0.8	4.8	4.8	0.05	4.8	0.10	0.19
24	40	6	0.1	4.8	2.8	0.8	0.01	0.01	0.05	4.8	1.64	1.56
25	40	0.1	6	4.8	2.8	0.01	0.01	0.01	19	0.01	0.20	0.23

A definitive screening design was developed to assess the impact of 10 ingredients found within the chemically defined media. All concentrations are in g/L except trace elements, amino acids and vitamins which are mL/L. Trace element working concentration contained (g/L): 15 $FeSO_4.7H_2O$, 2.4 $MnSO_4.H_2O$, 2.4 $ZnSO_4.7H_2O$, and 0.48 $CuSO_4.5H_2O$. A 100x stock amino acid mix contained (g/L): 12.9 arginine, and 10 each of histidine, leucine and methionine. A 1000x vitamin stock contained (g/L): 0.1 pyridoxine, 0.02 folic acid, 0.05 each of thiamine, riboflavin, niacin, pantothenic acid and nicotinamide. Abbreviations: Fru., fructose; T.E. trace element mixture; A.A., amino acid mixture; Vit., vitamin mixture. The DSD was performed in n = 2 biological replicates.

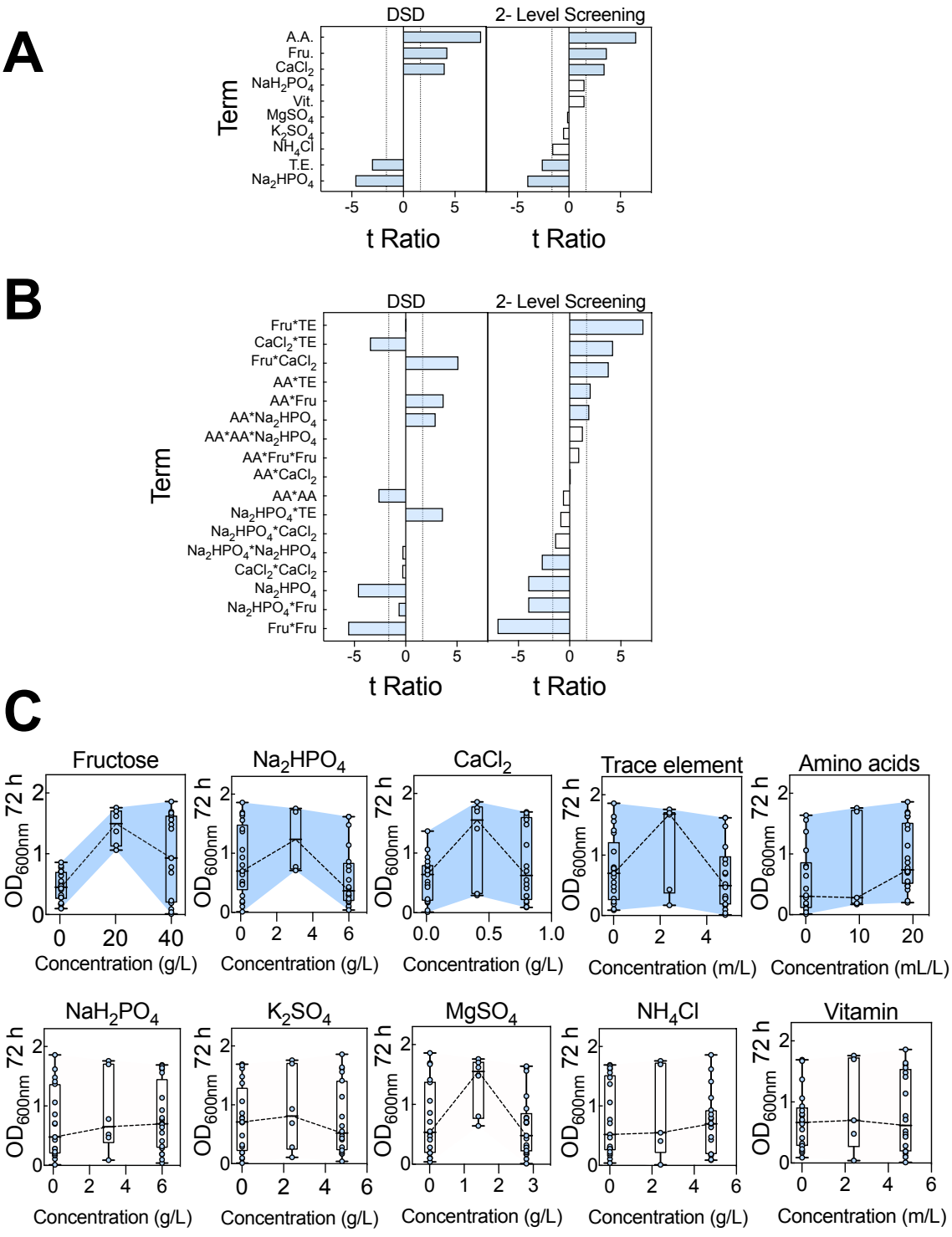


Fig. 2. 3 Definitive screening design array analysis.

Definitive and 2-Level screenings of data from DSD1 were performed. **A.** Main effect of each component. **B.** Two-way interaction between components. The comparative lengths of the t-ratios for each factor and factor interaction are shown. Bars extending to the right have a positive impact on growth at high concentration, those extending to the left have a negative impact on growth at high concentration. The broken vertical lines indicate threshold (t < -1.65 or > 1.65) level at 90% confidence level; term (factor or interaction) surpassing this threshold is deemed to have significant effect on the output, OD_{600nm}. Terms deemed significant for model projection are highlighted (blue). Analysis is based on two biological replicated arrays (Table 2.3). **C.** Summary of main effect plot from for each factor. The DSD was performed as n = 2 biological replicates.

2.3.3 Augmentation of data set based on significant growth components

A definitive screening design can force many of the data-points collected to the edges of the design space. The initial DSD provided evidence that the extreme concentrations of some of the components in the media were detrimental to growth. Therefore, the concentration ranges for the next stage of data collection could be more conservative than during the initial DSD. For example, fructose at low (0.5 g/L) or high (40 g/L) concentrations resulted in less and more variable growth over 72 h suggesting that maintaining a fructose concentration above 0.5 and below 40 g/L is the key to establishing robust and reliable growth. Guided by the first definitive screening design (DSD1), a second DSD array (DSD2) was created (Table 2.4). In this new design, factors under investigation were restricted to those that were highlighted as statistically significant in the DSD1. Fructose was restricted to 5–25 g/L, whilst Na_2HPO_4, $CaCl_2$, trace elements and amino acids were restricted to: 0.1–3.05 g/L, 0.1–0.459 g/L, 0.01–2.4 mL/L, 5 and 20 mL/L, respectively. Other factors that had not been identified as statistically significant were kept at concentration that resulted in a better growth from the DSD1. Vitamin solution was excluded from subsequent experiments. This is based on the information obtained from the analyses. Inclusion of vitamins did not contribute significantly to growth at the concentrations tested (Fig. 2.3A). Moreover, vitamin solution did not interact with any component (Fig. 2.3B), and its exclusion from medium resulted in cell growth comparable to that of medium with vitamins (Fig.1, Appendix A).

Table 2. 4 Definitive screening design 2.

Media	Fru.	NaH_2PO_4	Na_2HPO_4	K_2SO_4	$MgSO_4$	$CaCl_2$	NH_4Cl	T.E.	A.A.	OD1	OD2
1	15	3.05	0.1	2.41	1.41	0.1	0.01	0.01	5	1.71	1.36
2	15	3.05	3.05	2.41	1.41	0.46	0.01	2.41	20	1.84	1.67
3	25	3.05	1.58	2.41	1.41	0.46	0.01	0.01	5	1.72	1.65
4	5	3.05	1.58	2.41	1.41	0.1	0.01	2.41	20	1.95	1.95
5	5	3.05	0.1	2.41	1.41	0.28	0.01	2.41	5	1.85	0.99
6	25	3.05	3.05	2.41	1.41	0.28	0.01	0.01	20	0.57	1.84
7	5	3.05	3.05	2.41	1.41	0.46	0.01	1.28	5	1.18	1.18
8	25	3.05	0.1	2.41	1.41	0.1	0.01	1.28	20	2.10	2.06
9	5	3.05	3.05	2.41	1.41	0.1	0.01	0.01	12.50	2.03	1.99
10	25	3.05	0.1	2.41	1.41	0.46	0.01	2.41	12.50	1.86	1.40
11	15	3.05	1.58	2.41	1.41	0.28	0.01	1.28	12.50	1.92	1.96
12	5	3.05	0.1	2.41	1.41	0.46	0.01	0.01	20	2.01	2.08
13	25	3.05	3.05	2.41	1.41	0.1	0.01	2.41	5	1.06	0.91

A second definitive screening design (DSD2) was performed to assess the impact of the five ingredients identified as contributing significantly to *C. necator* growth. All concentrations are in g/L except trace elements and amino acids, which are mL/L. Trace element working concentration contained (g/L): 15 $FeSO_4.7H_2O$, 2.4 $MnSO_4.H_2O$, 2.4 $ZnSO_4.7H_2O$, and 0.48 $CuSO_4.5H_2O$. A 100x stock amino acid mix contained (g/L): 12.9 arginine, and 10 each of histidine, leucine and methionine. Abbreviations: Fru., fructose; T.E., trace element mixture; A.A., amino acid mixture. The DSD2 array was performed in duplicate.

Examining the combined data for DSD1 and DSD2 did indeed confirm that the extreme concentrations of some of the components were detrimental to cell growth. For example, when the concentrations of fructose were at the highest and lowest values (0.5 and 40 g/L, respectively) the OD_{600nm} 72 h were both lower and more variable than when fructose was restricted to between 5–25 g/L (Fig. 2.4A). This indicates that maintaining fructose between 5 and 25 g/L is key to establishing robust and reliable growth. Adjusting $CaCl_2$ to 0.1–0.46 g/L resulted in robust and reliable growth compared with extreme concentrations from the DSD1 (Fig. 2.4B). Similarly, lower concentration of Na_2HPO_4 and trace element solution resulted in better growth (Fig. 2.4C and D). At this stage, it appears that—with the exception of amino acids—maintaining fructose, Na_2HPO_4, $CaCl_2$ and trace element at moderate concentrations resulted in robust growth. Growth increased with increasing concentration of amino acid (Fig. 2.4E). Maximum OD_{600nm} obtained from the DSD2 was 2.10 in comparison to 1.76 obtained from DSD1 indicating that adjusting the concentration of some components did indeed result in a better output.

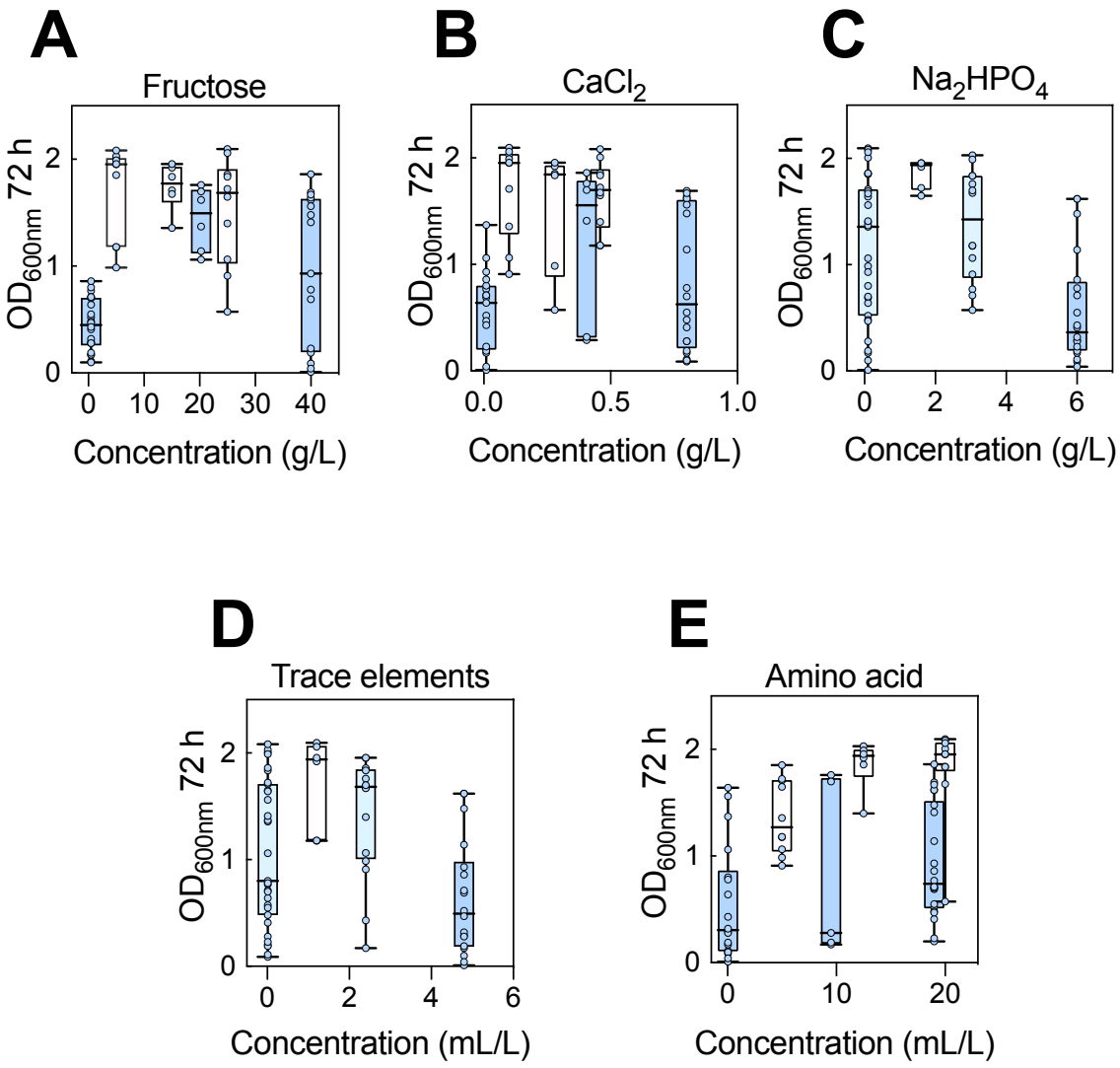


Fig. 2. 4 Combined data for DSD1 and DSD2 for key media components.

A. Fructose. **B.** $CaCl_2$. **C.** Na_2HPO_4. **D.** Trace elements. **E.** Amino acids. Data shown are for n = 2 biological replicated arrays of each definitive screening design: DSD1 (blue boxes) and DSD2 (white boxes). Settings were both DSDs were held at the same concentration are coloured teal. Error bars are for standard error mean (S.E.M.).

Considering the interaction highlighted between amino acids and trace elements in DSD1, it was further observed in DSD2 that high and reliable growth were obtained when both components were kept at high concentrations (20 mL/L and 2.405 mL/L, respectively), or when trace elements were kept at very low concentration (0.01 mL/L) while amino acid concentrations were cautiously reduced. The opposite scenarios, in which trace elements were at high concentrations and amino acids were kept low, or when both elements were kept low impacted negatively on growth, resulting in variable OD_{600nm} at 72 h (Table 2.4). In addition, for interactions between fructose and Na_2HPO_4, and between fructose and $CaCl_2$, it

was observed that (assuming balanced amino acid and trace elements were maintained), a high fructose concentration (> 20 g/L) is compensated with low Na_2HPO_4 (< 1.58 g/L) and high calcium concentrations (> 0.1 g/L). Nevertheless, low fructose concentration (5 g/L) is mainly compensated with low $CaCl_2$ (≤ 0.1 g/L) regardless Na_2HPO_4 concentration—provided amino acid and trace concentrations are maintained at concentrations permitting high growth. Furthermore, for amino acid and fructose interaction, high amino acid concentration (> 10 mL/L) supported robust growth and low concentration impacted negatively on growth. The negative impact of low amino acid concentration (≤ 5 mL/L) appeared to be greater under high fructose concentration. It follows that at such low amino acid concentration, adjusting fructose concentration to 5 g/L will likely support high growth, with little probability. Although formulations containing high concentration of amino acid (> 10 mL/L) resulted in higher and more reproducible growth, such high concentration contributed ~ 12% to growth in the absence of fructose as a carbon source, whereas under the same formulation, concentration ≤ 10 mL/L amino acid contributed < 5% to growth. Hence, in subsequent experiments, amino acid concentration was set at 10 mL/L and was compensated with ≤ 1 mL/L trace. Moreover, this information allowed establishment of a target zone for growth in which an OD_{600nm} of between 1.5 and 2.2 is reached at 72 h.

2.3.4 Identifying the ingredient(s) responsible for amino acid and trace interaction

With the augmented data (DSD1 and DSD2) it became clear that there was an interaction between amino acids and trace elements. More so, given that both components are potential growth supplements and that each contained mixture of other ingredients, it was necessary to investigate the interaction to determine which ingredients that are responsible for the observed interaction between the two components. To do this, a medium that permits robust cell growth over 72 h was formulated and used as a control.

It was observed that the absence of trace elements did not adversely affect growth of *C. necator* but that the absence of amino acids did (Fig. 2.5A). Interestingly, simultaneous exclusion of both amino acids and trace elements resulted in cell densities comparable to the control. Therefore, it was hypothesised that in the absence of one or more of the amino acids (methionine, histidine, leucine and/or arginine), one or more of the components of the trace elements ($CuSO_4$, $FeSO_4$, $MnSO_4$ and/or $ZnSO_4$) inhibits growth of *C. necator*. The hypothesis was tested first by formulating media containing no amino acids and withdrawing each trace element in turn (Fig. 2.5B). Under these conditions, media without $CuSO_4$ but containing other trace elements resulted in growth comparable to the control, while all three formulations that contained $CuSO_4$ had reduced cell densities. These observations support

the hypothesis that a component in trace solution, in the absence of amino acids, inhibits *C. necator* growth. This component was established to be $CuSO_4$ (Fig. 2.5B).

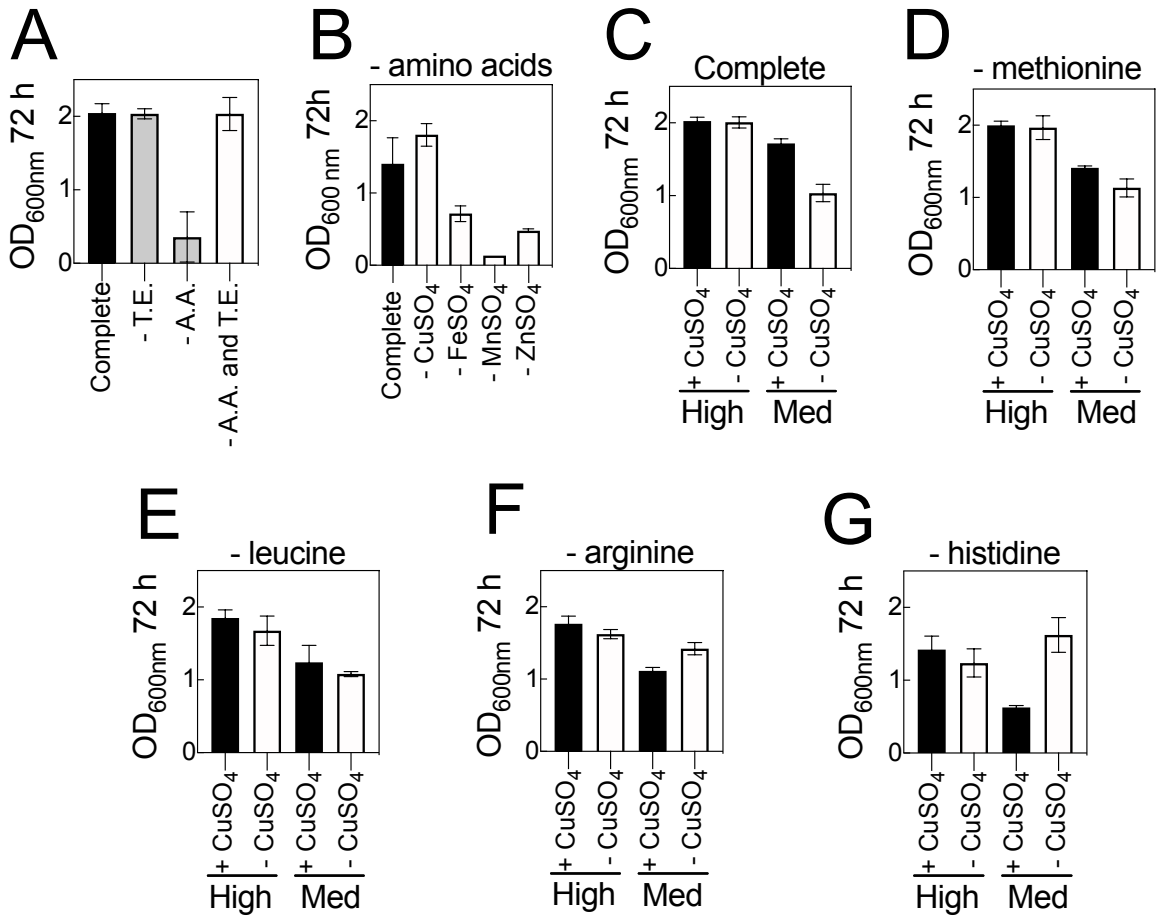


Fig. 2. 5 Interactions between trace elements and amino acids.

A. *C. necator* grown in a complete chemically defined medium with amino acids, trace elements or both amino acids and trace elements excluded. **B.** *C. necator* grown in a complete medium with each of the four trace elements excluded. **C–G.** *C. necator* grown in a complete chemically defined medium with or without $CuSO_4$ and with each of the four amino acids excluded. High: 20 mL/L amino acid; Med (medium): 10 mL/L amino acid. Each trace was added at 2.4 mL/L; the working concentration of each trace element solution is as shown in Table 2.3. Error bars are S.E.M., $n = 2$.

To further determine which amino acid(s) interacts with $CuSO_4$, a series of experiments were performed in which each amino acid was excluded in formulations with and without $CuSO_4$. When $CuSO_4$ was included in media the concentration was kept constant at 2.405 mL/L, the concentration used in testing the previous hypothesis. Cognisant that high concentration of amino acids will contribute more to growth, potentially serving as a carbon source, the experiment was performed at two amino acid concentrations: med (10 mL/L) and high (20 mL/L).

As expected, at high concentrations of amino acids the presence or absence of $CuSO_4$ did not affect growth (Fig. 2.5C). However, at medium concentrations of amino acids growth was partially suppressed in the presence of $CuSO_4$ but that this was exacerbated in the absence of $CuSO_4$. Therefore, $CuSO_4$ is considered an important medium component and cannot simply be excluded from formulations. Similar growth responses were seen in experiments in which either methionine or leucine were excluded (Fig. 2.5D and E). Under medium amino acid concentration, if arginine, or most notably histidine, were removed from media, then the presence of $CuSO_4$ impaired growth of *C. necator*, resulting in < 1.0 OD_{600nm} (Fig 2.5F and G). For media lacking histidine this effect was also observable when all other amino acid levels were maintained at high concentrations (Fig. 2.5G). Interestingly, growth was restored when both histidine and $CuSO_4$ were excluded under medium concentration of other amino acids. However, the level of growth (~ 1.8 OD_{600nm}) was slightly lower in comparison to that of complete medium (> 2.0 OD_{600nm}). From this data it was deduced: first, that histidine protects against the inhibitory effects of copper, but secondly, $CuSO_4$ is an important component of the media whose absence retards growth when amino acid content, with the exception of histidine, is restricted. Both $CuSO_4$ and histidine concentrations must therefore be balanced for robust growth. It was therefore established that histidine and $CuSO_4$ are the main ingredients responsible for the observed interaction between the amino acid solution and trace elements. Thus, the interaction between $CuSO_4$ and histidine is a concentration dependent positive interaction. This resulted in high and reliable growth when both ingredients are present. Hence, in subsequent experiments, both $CuSO_4$ and histidine were included in media formulations.

2.3.5 Confirming the impact of NaH_2PO_4, K_2SO_4, $MgSO_4$ and NH_4Cl on growth

With a greater understanding of which components and concentrations are required to formulate a medium supporting robust growth, components (NaH_2PO_4, K_2SO_4, $MgSO_4$ and NH_4Cl) that were identified by the DSD1 as not statistically significant were re-investigated for their effect on growth. In the original DSD, these factors were not identified as being statistically significant, but those experiments were run under conditions in which key components (e.g. fructose and amino acid) were at concentrations that have since been identified as resulting in poor or unreliable growth responses. The non-significant components were then re-investigated under conditions in which Na_2HPO_4, $CaCl_2$, trace elements and amino acid concentrations were not disruptive to cell growth. Fructose was set at either 5 g/L or 20 g/L (Table 2.5); this was to determine if concentration of carbon source has an impact on main effect of the components under investigation.

Table 2. 5 Definitive Screening Design 3.

Media	Fru.	NaH_2PO_4	Na_2HPO_4	K_2SO_4	$MgSO_4$	$CaCl_2$	NH_4Cl	T.E.	A.A.	OD1	OD2
1	5	0.1	0.1	0.01	0.01	0.46	4.8	1	10	0.13	0.17
2	5	0.1	0.1	1.21	1.41	0.46	0.01	1	10	1.79	1.74
3	5	0.1	0.1	2.41	0.01	0.46	0.01	1	10	1.64	1.71
4	5	0.1	0.1	0.01	1.41	0.46	2.41	1	10	0.95	0.57
5	5	3.05	0.1	1.21	0.01	0.46	4.8	1	10	1.85	1.76
6	5	3.05	0.1	0.01	0.71	0.46	0.01	1	10	1.86	1.65
7	5	3.05	0.1	2.41	1.41	0.46	0.01	1	10	1.89	1.88
8	5	1.56	0.1	0.01	0.01	0.46	0.01	1	10	1.76	1.92
9	5	3.05	0.1	0.01	1.41	0.46	4.8	1	10	0.84	0.79
10	5	0.1	0.1	2.41	0.71	0.46	4.8	1	10	0.66	0.73
11	5	1.56	0.1	2.41	1.41	0.46	4.8	1	10	1.51	1.55
12	5	3.05	0.1	2.41	0.01	0.46	2.41	1	10	1.95	1.89
13	5	1.56	0.1	1.21	0.71	0.46	2.41	1	10	1.49	1.76
14	20	0.1	0.1	0.01	0.01	0.46	4.8	1	10	0.15	0.81
15	20	0.1	0.1	1.21	1.41	0.46	0.01	1	10	1.86	2.09
16	20	0.1	0.1	2.41	0.01	0.46	0.01	1	10	2.09	1.92
17	20	0.1	0.1	0.01	1.41	0.46	2.41	1	10	1.66	0.95
18	20	3.05	0.1	1.21	0.01	0.46	4.8	1	10	1.79	1.52
19	20	3.05	0.1	0.01	0.71	0.46	0.01	1	10	1.71	1.90
20	20	3.05	0.1	2.41	1.41	0.46	0.01	1	10	2.05	1.73
21	20	1.56	0.1	0.01	0.01	0.46	0.01	1	10	1.81	1.85
22	20	3.05	0.1	0.01	1.41	0.46	4.8	1	10	1.08	0.98
23	20	0.1	0.1	2.41	0.71	0.46	4.8	1	10	1.17	0.96
24	20	1.56	0.1	2.41	1.41	0.46	4.8	1	10	1.65	1.48
25	20	3.05	0.1	2.41	0.01	0.46	2.41	1	10	1.74	1.80
26	20	1.56	0.1	1.21	0.71	0.46	2.41	1	10	1.65	1.39

A final set of experiments were performed to re-assess the impact of components not deemed significant in the first DSD. All concentrations are in g/L except trace elements, amino acids and vitamins, which are mL/L. Trace element working concentration contained (g/L): 15 $FeSO_4.7H_2O$, 2.4 $MnSO_4.H_2O$, 2.4 $ZnSO_4.7H_2O$, and 0.48 $CuSO_4.5H_2O$. A 100x stock amino acid mix contained (g/L): 12.9 arginine, and 10 each of histidine, leucine and methionine. Abbreviations: Fru., fructose; T.E., trace element mixture; A.A., amino acid mixture. The DSD array was performed in duplicate.

The results indicated that these components did indeed affect growth rate when primary factors are not restricting growth (Fig. 2.6). Monosodium phosphate (NaH_2PO_4) and K_2SO_4 had some detrimental impact on growth when concentrations were low (Fig. 2.6A and B), whilst varying the concentration of $MgSO_4$ had no significant effect on growth (Fig. 2.6C). Some of the suggestions from the DSD1 were confirmed. Most notable was the observation that increasing NH_4Cl concentration had a detrimental effect on culture cell density (Fig. 2.6D). These results were true at both high and low concentrations of fructose. Formulations having high concentration of NH_4Cl (> 2.41 g/L) resulted in low growth; while those having it at low concentration (0.01 g/L) resulted in high growth comparable to that of the positive control formulation. However, the negative impact of high NH_4Cl concentration appeared to be nullified when NaH_2PO_4 and/or K_2SO_4 were kept at high concentrations (Table 2.5).

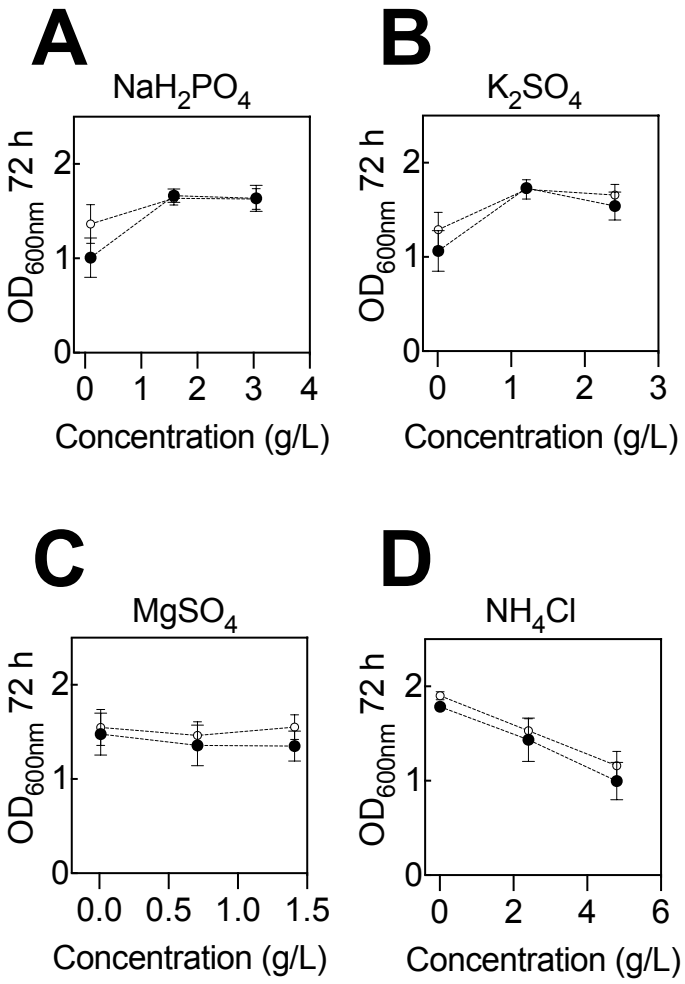


Fig. 2. 6 Re-examination of non-significant media components.

The impact of components not deemed statistically significant in DSD1 were re-examined under less extreme conditions. These were: **A.** NaH_2PO_4. **B.** K_2SO_4. **C.** $MgSO_4$. **D.** NH_4Cl. Experiments were conducted at 5 g/L (open circles) or 20 g/L (closed circles) fructose. Error bars are S.E.M., n = 6.

2.3.6 Modelling and validation of the media formula-growth response landscape

At this stage, 64 different variant formulations have been experimentally assessed in duplicate across three different experiments (25, 13 and 26 experimental runs respectively). Based on this, a statistical model trained against this data set, that could describe the understanding of how the cell cultures respond to changes in media composition and predict performance in novel formulations was built. Two-level screening was performed on all 128 runs. This identified a few factors and factor interactions deemed significant for model projection, as earlier highlighted. However, this screening (analysis) involving all data did not highlight fructose or K_2SO_4 (as these had been restricted to concentrations that did not significantly impact growth during much of DSD2 and DSD3). Nevertheless, these terms were included manually in the model, as they had been previously highlighted as significant factors. These terms were used to construct a standard least squares model. The least squares model was able to describe the relationships within the data with good accuracy (Fig. 2.7A). Although the model was internally consistent, it was important to know if it could predict OD_{600nm} at 72 h in formulations it had not encountered during model training. To validate the model prediction, 16 new formulations with sampling biased towards media formulations that were predicted to be in the top 25% of media performance were composed. Each of these was assessed in triplicate and the resulting OD_{600nm} compared to predicted OD_{600nm} (Fig. 2.7B and C). As predicted by the model, all of the new formulations fell within the upper quartile of formula performance (Fig. 2.7B). More importantly, the experimental OD_{600nm} correlate well with the model predicted OD_{600nm} (Fig. 2.7C).

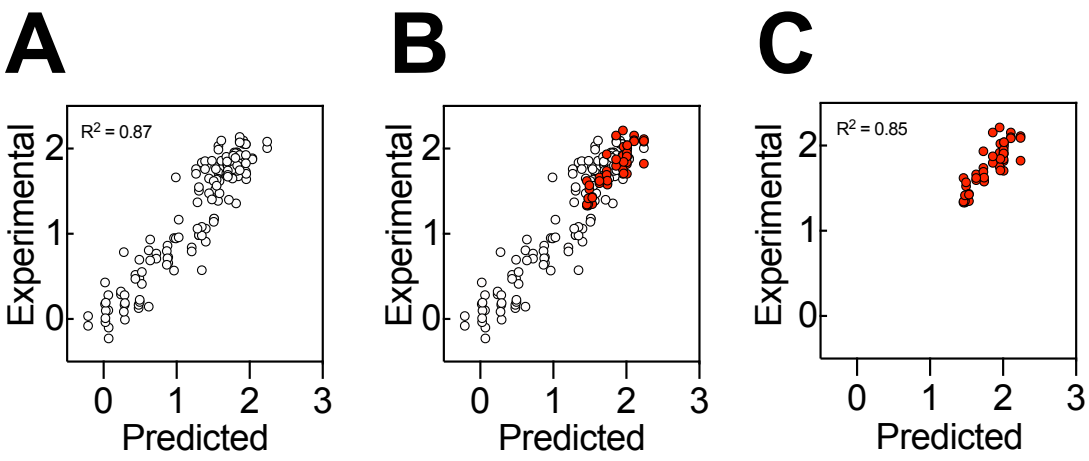


Fig. 2. 7 Experimental validation of model predictions in 48-well plate format.

A. Model predicted values plotted against experimental data for least squares model. **B.** New experimental data for 1 mL culture volumes overlaid against the original least squares model. Open circles show the training data set, red circles are new data. **C.** New experimental data for 1 mL cultures plotted on their own against the model predicted values. Three biological replicates were assessed for each prediction.

2.3.7 Description of model projections and growth predictions

The model allowed us to visualise phenomenon observed during the data-collection phase of the investigation. For example, it visualises the interaction between amino acids (specifically histidine) and trace elements (specifically copper) that was elucidated in Fig 2.5. It shows that high concentrations of trace elements are detrimental to growth and that increasing amino acid concentration can help mitigate the inhibitory effects of high concentrations of trace elements (Fig. 2.8A). The greater the concentration of trace elements included, the higher the amino acid concentration needs to be. Nevertheless, increasing the concentration of amino acids impacts in other areas. A strong monotonic relationship was observed between amino acid and fructose. Increasing the concentration of both components resulted in increased growth. Although lower concentration of amino acid (≤ 5 mL/L) is not considered safe for reproducible growth, the negative impact of the lower concentration on growth appears to be greater under high fructose concentration (Fig. 2.8B). It follows that at 5 mL/L of amino acid, adjusting fructose concentration to 5 g/L will likely support better growth. Given that the amino acids present in the media are the only alternative source of carbon for growth, raising the concentration of amino acids too high will impact on interpretation of experiments designed to examine carbon utilisation. A balance therefore needs to be struck between fructose, amino acids and trace element concentrations. The model also visualises interaction between Na_2HPO_4 and fructose (Fig. 2.8C). Greater concentrations of Na_2HPO_4 results in lower cell densities—an effect that can be partially offset by decreasing fructose concentrations. The model also suggests that if $CaCl_2$ is to be increased, fructose concentration must be increased and vice versa (Fig. 2.8D). Lastly, the negative effect of increasing NH_4Cl concentrations had been previously observed (Fig. 2.6D). The model indicates that this negative effect can be nullified by increasing the concentration of K_2SO_4 and/or NaH_2PO_4 (Fig. 2.8E and F). Understanding the interactions between media components is vital for predictions of culture performance and allows changes to be made to media formulations for different experimental goals.

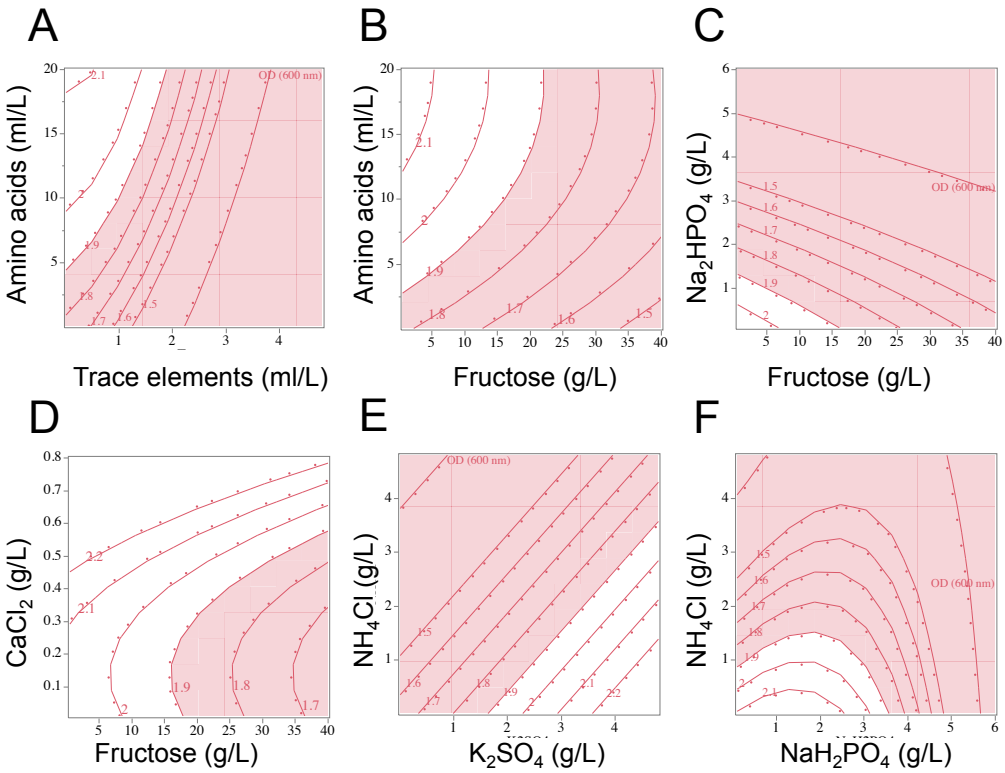


Fig. 2. 8 Interactions of components of the defined media described by the least square model.

A. Amino acid and trace element interactions. **B.** Fructose and amino acid interactions. **C.** Disodium phosphate and fructose interactions. **D.** Calcium chloride and fructose interactions. **E.** Ammonium chloride and potassium sulphate. **F.** Ammonium chloride and monosodium phosphate. Each panel represents a two way-interaction. Red (pink) zone is an area where OD_{600nm} at 72 h fails to reach 1.9, and white zone is where it surpasses 1.9. Each contour grid line represents an OD_{600nm} of 0.1 increment. All other media components were kept at concentrations permitting high growth.

2.3.8 Model validation at greater volumes

Having built and experimentally validated the growth model at small-scale cultivation (1 mL), 48-well plates, the model was further tested for its ability to predict *C. necator* growth at larger-scale cultivations (100 mL and 1 L). Shake flask cultivations were carried out under identical conditions (30°C, 200 rpm) over 72 h periods in 250 mL baffled and non-baffled flasks, each containing 100 mL of medium. Each cultivation formula was randomly selected from L32 fractional factorial design of experiment (Fig. 2.9). Optical densities (OD_{600nm}) for each formula in both types of flasks were strikingly similar at every interval throughout the cultivation period. Most importantly, the growth ranks for baffled and non-baffled flask cultivations were identical. Spearman's correlation showed a significant ($p < 0.05$) relationship between predicted and actual ranks for both flask types; the correlation coefficient was $\rho = 0.87$ (Fig. 2.10A).

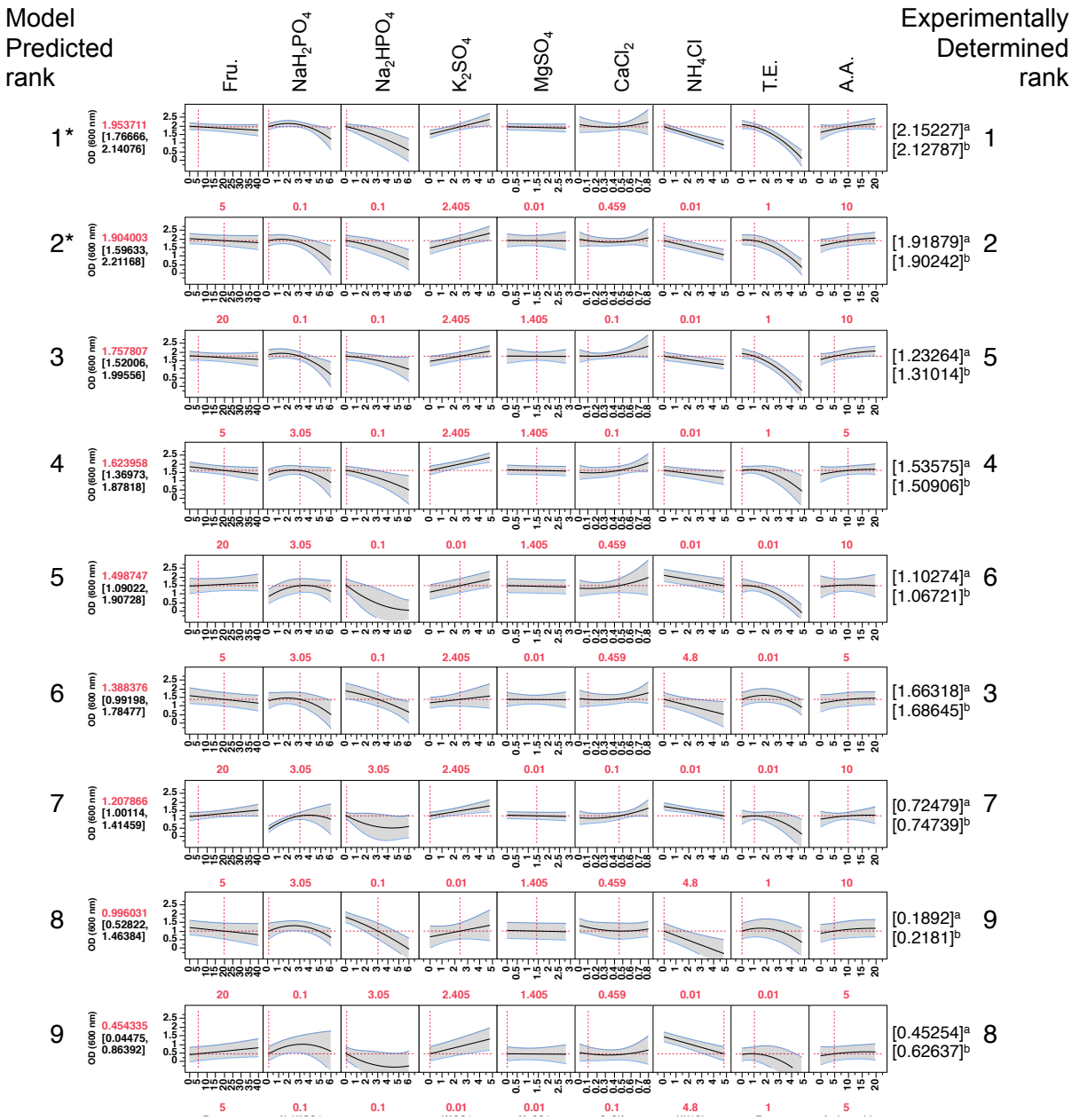


Fig. 2. 9 One hundred millilitre validation. Nine different formulations were assessed at 100 mL culture volumes.

For each row, the model predicted rank and OD_{600nm} 72 h (values in brackets) are indicated on the left, media settings for each component are indicated in red underneath, the measured OD_{600nm} 72 h for two cultivation experiments (values in brackets) and experimentally determined rank are indicated on the right. Non-baffled flask cultivations (a); baffled flask cultivations (b); included in bioreactor cultivations and subsequently served as controls (*). All concentrations are in g/L except T.E. and A.A, which are in mL/L. Abbreviations: Fru., fructose; T.E., trace element mixture; A.A., amino acid mixture. The output is *n* = 2 biological replicated arrays of definitive screening designs (DSDs).

Next, two formulae from shake flask cultivation together with three additional formulae were cultivated in 1 L bioreactor (Fig. 2.11). Similarly, growth rank was predictable, with significant relationship between predicted rank and actual growth rank (ρ = 0.90) (Fig. 2.10B). During the cultivations, there were no significant changes in bioprocess parameters. Constant agitation at 200 rpm, with 1 vvm airflow rate was sufficient to maintain dissolved oxygen (dO_2) above 20%. Although no base was added in all cultivations, the pH of the media did not drop below 4.5, the set point. In addition, there was no correlation between pH of media (ρ = 0.23 and ρ = 0.52 for shake flask cultivation and bioreactor cultivation, respectively) prior to inoculation or after growth at 72 h (Fig. 2, Appendix A).

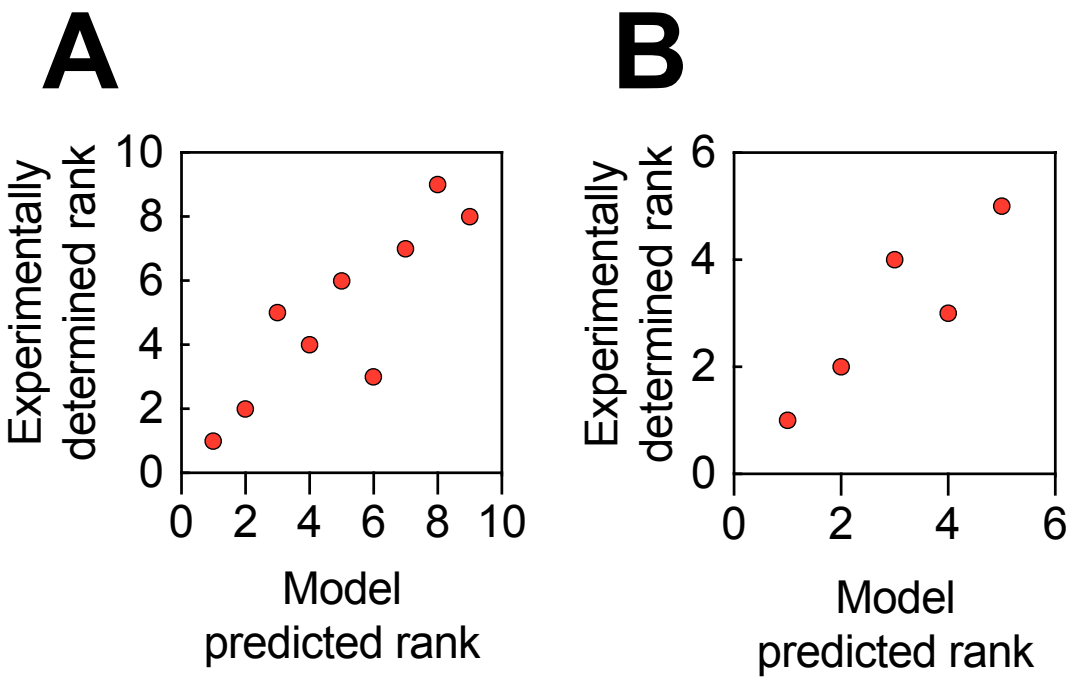


Fig. 2. 10 Experimental validation of model predictions at shake-flask and bioreactor scale.

A. Spearman's rank correlations (ρ = 0.87) between model predicted rank and experimental data for 100 mL culture volumes. **B.** Spearman's rank correlations (ρ = 0.90) between model predicted rank and experimental data for 1 L bioreactor batch flask cultures.

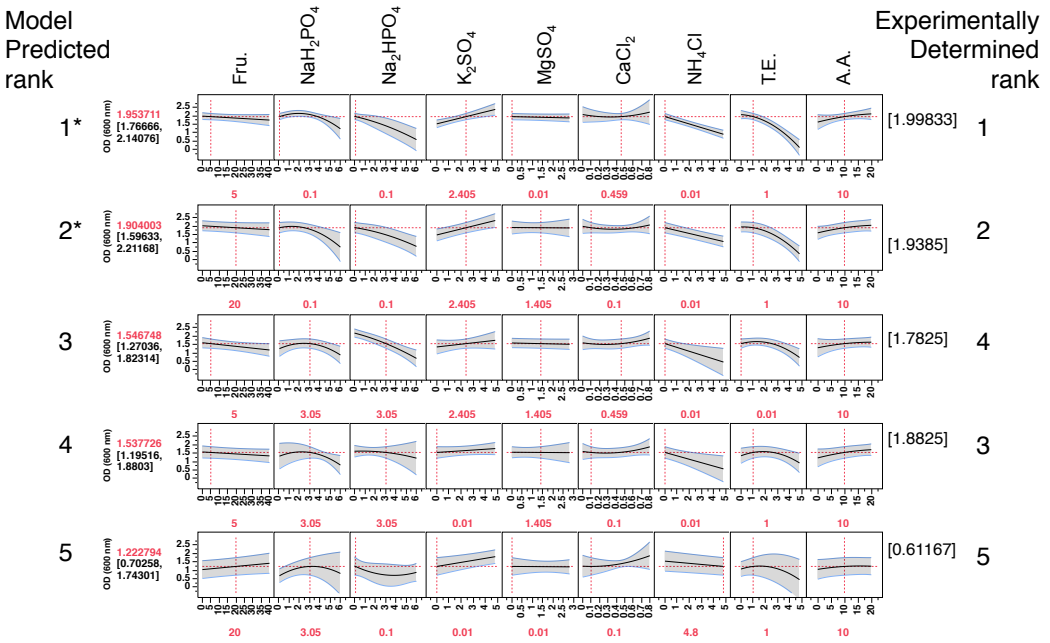


Fig. 2. 11 One litre validation. Five different formulations were assessed at 1 L culture volumes.

For each row, the model predicted rank and OD_{600nm} 72 h (values in brackets) are indicated on the left, media settings for each component are indicated in red underneath, the measured OD_{600nm} 72 h (values in brackets) and experimentally determined rank are indicated on the right. Controls (*). All concentrations are in g/L with exception of T.E. and A.A, which are in mL/L. Abbreviations: Fru., fructose; T.E., trace element mixture; A.A., amino acid mixture.

2.3.9 Distinguishing between statistically significant and essential media components

The growth model highlighted components and interactions that contribute significantly to growth. However, it is important to draw a distinction between what is statistically significant and what is essential. Determining the essential nature of components involves the conventional approach of withdrawing one factor individually during media composition. To ensure that the concentrations at which factors are held while individually withdrawing a component did not introduce bias on the essential nature of any component, two media were used as controls. One of these controls had every component maintained at low concentration (low concentration formula) while the other had every component maintained at high concentration (high concentration formula). Both controls were predicted to support robust growth of *C. necator* (Fig. 2.12). When each component was individually withdrawn in both formulae, similar growths were observed indicating that the concentration at which components were held did not impact on the essential nature of any components of the media (Fig. 2.13A and B). Little growths were observed in media lacking fructose or $MgSO_4$ whereas growths were significantly impaired in media lacking amino acid or K_2SO_4. The remainder components when individually withdrawn did not significantly impact on growth.

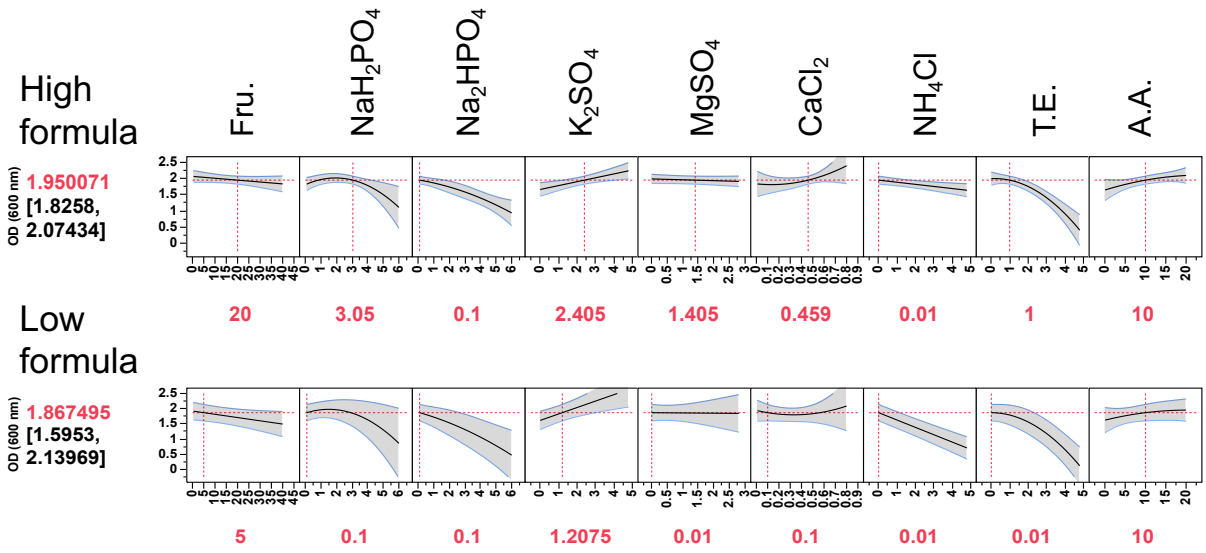


Fig. 2. 12 Prediction profiler of media at high and low composition.

The model predicted rank and OD_{600nm} 72 h (values in brackets) are shown on the left, media settings for each component are shown in red underneath. All concentrations are in g/L except of T.E. and A.A, which are in mL/L. Abbreviations: Fru., fructose; T.E., Trace element mixture; A.A., amino acid mixture. The output is *n* = 2 biological replicated arrays of definitive screening designs (DSDs).

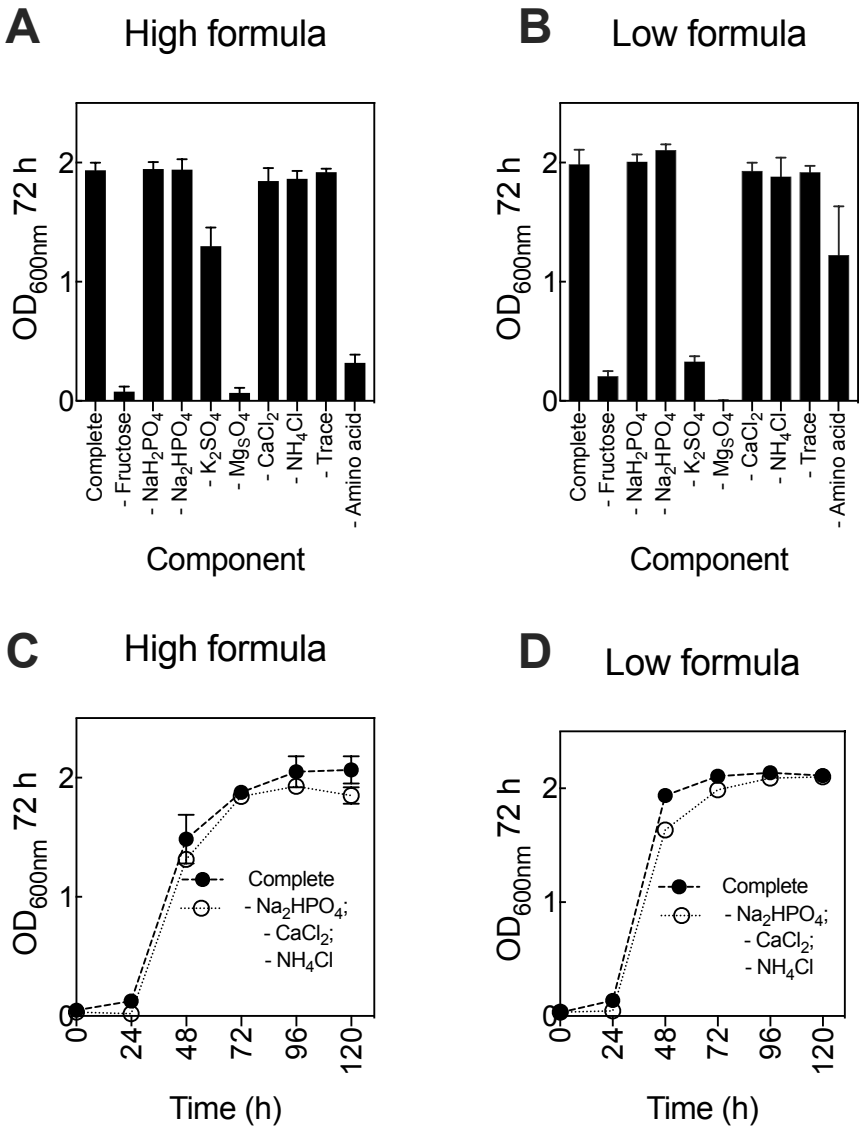


Fig. 2. 13 Determination of essentiality of each component in *C. necator* defined medium.

The impact of individually withdrawing each component on growth at: **A.** High concentration media formulation. **B.** Low concentration media formulation. The impact of simultaneously withdrawing some components deemed non-essential on growth at: **C.** High concentration media formulation; **D.** Low concentration media formulation. The controls (complete) were predicted to support robust growth at low and high concentration media formulations (Fig. 2.12). Error bars are S.E.M., *n* = 3 biological replicates.

Therefore, it is established that fructose and $MgSO_4$ are essential for cultivation of *C. necator* H16; yet $MgSO_4$ is not deemed statistically significant by the model. This is because under our experimental conditions, all concentrations of $MgSO_4$ tested are likely in excess and do not limit growth. There is therefore scope to reduce $MgSO_4$ concentrations if this was desirable. Based on the levels of growth, amino acid and K_2SO_4 are considered important, while the remainder components are considered non-essential for growth. Interestingly, for both formulae, simultaneously withdrawing some non-essential components (Na_2HPO_4,

$CaCl_2$, and NH_4Cl) resulted in similar growth pattern and gave OD_{600nm} comparable to that of the controls (Fig. 2.13C and D).

Finally, colony-forming unit per millilitre (cfu/mL) was correlated against OD_{600nm} using control media predicted to support robust growth. This was performed to further confirm that OD_{600nm} is indeed appropriate for measuring *C. necator* growth in defined media and that the OD_{600nm} observed throughout this study were ascribed to growth (viable cells) and not as a result of precipitation of components in media. Optical density (OD_{600nm}) and cfu/mL showed similar growth trends over 120 h cultivation (Fig. 2.14).

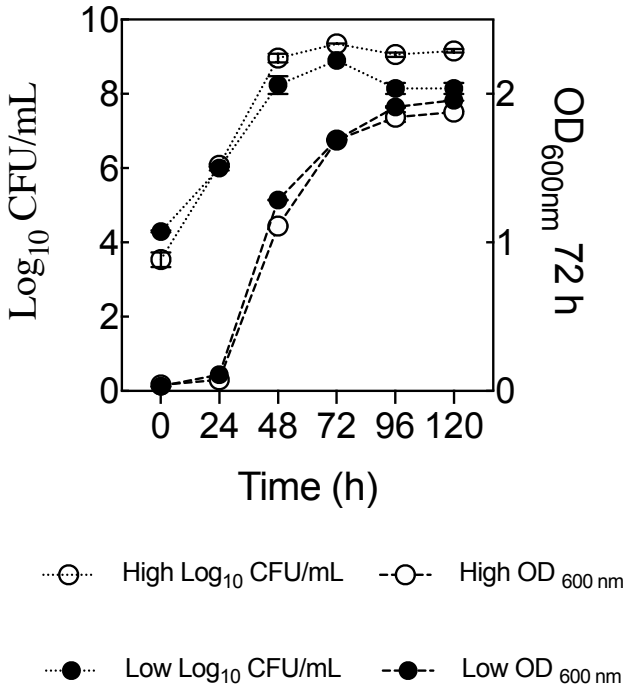


Fig. 2. 14 Validation of OD_{600nm} as surrogate for growth measurement.

Viable cells (cfu/mL) of *C. necator* H16 correlated with optical density (OD_{600nm}) over prolonged cultivation. Experiment was performed at low and high concentration media formulations guided by the growth model prediction (Fig. 2.12). Cell were grown on defined media and plated on LB agar and incubated as appropriate for *C. necator*. Error bars are S.E.M., n = 3 biological replicates.

2.4 Discussion

We developed a model trained against a structured set of data for the cultivation of *C. necator* H16 in a chemically defined medium with fructose as the primary carbon source. The systematic approach identified significant growth factors and their effects on culture density at 72 h. Fructose, glucose, glycerol and sucrose were used in the preliminary phase with fructose considered as the best carbon source supporting robust growth under heterotrophic conditions. While *C. necator* H16 has been reported to have broad substrate range, its ability to utilise carbohydrates as a carbon source during heterotrophic growth appears to be limited to fructose and *N*-acetylglucosamine [9,10,106]. Utilisation of fructose by *C. necator* is most likely to occur via substrate importation by an ATP-binding cassette (ABC-type) transporter, followed by catabolism via 2-keto-3-deoxy-6-phosphogluconate (KDPG), the Entner-Doudoroff pathway. The responsible genes, notably a putative regulator (*frcR*), ribosome transporters (i.e. *frcA*, *frcC* and *frcB* orthologs in *Escherichia coli* and *Ralstonia solanacearum*) and other essential genes facilitating such metabolism are located on chromosome 2 inside gene clusters for glucose, 2-ketogluconate, and glucosamine catabolism [9]. In contrast, phosphofructokinase and 6-phosphogluconate dehydrogenase (key enzymes of the Embden-Meyerhof-Parnas and the oxidative pentose phosphate pathways, respectively) appear to be absent from the *C. necator* H16 genome. It is perhaps not surprising that glucose supported little or no growth, yet it has been reported that prolonged cultivation (> 70 h) with glucose as the sole carbon source resulted in a mutant that was able to utilise glucose [107]. The mutant acquired such ability by mutating the *N*-acetylglucosamine phosphotransferase system, which when deleted led to inability to utilise glucose [107,108].

Glycerol supported low growth of *C. necator* especially in the medium forcing during the scoping trial. Such low growth was attributed to oxidative stress resulting from high levels of reactive oxygen species (ROS) formed as a result of elevated level of hydrogenases, leading to cell destruction by damaging DNA and other cell components [36]. The elevated levels of hydrogenases, soluble and particulate, were due to limited availability of energy [109]. When *C. necator* was cultivated under heterotrophic condition with glycerol as the carbon source, proteins (bifunctional catalase/peroxidase, oxygen-liable aconitase, peroxiredoxin, hydroperoxidase and glutathione peroxidase) involved in ROS destruction were up regulated in comparison to that observed with another carbon source, succinate [36]. Interestingly, *C. necator* growth on glycerol as the primary carbon source appears not to be entirely heterotrophic due to high predicted flux of Calvin-Benson-Bassham (CBB) cycle enzymes compared to the flux observed with another carbon source such as fructose [110]. The

confirmation of such active CBB cycle [110], together with high level of hydrogenases [109] indicates that *C. necator* growth with glycerol is grafted for mixotrophic rather than heterotrophic growth [111].

Formulations of chemically defined media previously described for the cultivation of *C. necator* H16 are typically prepared without amino acids [11,13,16,20,44,92]. In this study, *C. necator* growth is improved when a small number of amino acids (arginine, histidine, leucine and methionine) are included in media. The data suggest that they serve as preferred sources of nitrogen, compared to NH_4Cl. Their inclusion resulted in a faster growth rate compared to media lacking amino acids. Similar observation was also made when these amino acids were included in the medium that served as a guide for choice of components [44]. Although the effect of histidine is greater than that of methionine, arginine and leucine, their action was synergistic. These amino acids are essential for cultivation of certain bacterial genus [112]. Interestingly, the model indicated that there was an interaction between fructose and amino acid content that indicates the carbon:nitrogen ratio balance must be maintained, irrespective of the actual values. Growth reductions under low amino acid concentration (5 mL/L) are less severe in media containing low concentration of fructose (5 g/L) than in media with high fructose concentration (> 20 g/L).

Metal ions, especially divalent cations of d-block transition metals, are important for bacterial growth where they act as metalloenzyme cofactors in living cells. Their presence in high concentrations is detrimental to cells [113]. Copper is known to play diverse structural and catalytic functions owing to its ability to exist either in a reduced (Cu^+) state, with affinity for thiol and thioether groups, or in an oxidized (Cu^{2+}) state with more likely coordination for oxygen or imidazole nitrogen group of amino acids including histidine [114,115]. Bacteria have been reported to respond to high levels of copper by one of three families of metalloregulatory repressors: CopY, CsoR, and CueR. Under high concentration, an integral membrane Cu^+ transporter (P_{1B}-type ATPase)—characterised by histidine-rich domains, and conserved cytosine/histidine motifs within specific transmembrane domains—exports copper from the cytoplasm into the periplasm where in Gram-negative bacteria further detoxification and exportation are carried out by other related enzymes such as Cu^+ oxidase, Cu chaperones and cation transporters [114,115]. Further, studies have reported that copper toxicity affect sugar (glucose) utilisation in different microorganisms, and its growth inhibition on fungi (*Neurospora crassa* and *Saccharomyces cerevisiae*) was due to impaired/suppression of histidine biosynthesis [116,117]. In general, the toxic effect of copper appears to be enhanced under histidine limitation or impaired histidine biosynthesis and is neutralised by adding

histidine. Further details on the role of histidine in diminishing toxicity of excess copper is discussed Chapter 6.2.

Media containing high concentration of NH_4Cl had similar pH values comparable to those containing low concentration. Thus, the detrimental effect of high concentration of NH_4Cl on growth was not as a result of high (alkaline) pH, rather it was ascribed to enhanced osmolarity or medium ionic strength [118]. High concentration of ammonium is not specifically toxic to different groups of bacteria (Gram-positive and -negative) owing to their ability to diffuse it across cytoplasmic membrane into cell, or back into medium. Importantly, these groups of bacteria tightly regulate internal ammonium concentration by synthesising special ammonium transporters only under nitrogen limitation [118–120].

Although the importance of calcium ion on bacterial growth is not fully understood, studies have suggested it acts as a regulator to maintain protein stability, enzyme activity or signal transduction [121–123]. Phosphorus is amongst the three major macro-elements required for biological activities. Besides serving as a source of phosphate, NaH_2PO_4 and Na_2HPO_4 also act as buffering agents to maintain pH of media. The two-way interaction between fructose and Na_2HPO_4 is attributed to decrease diffusion of hydrogen phosphate under high fructose concentration [124]. The most abundant cation in bacterial cells, potassium ion (K^+), is considered an important nutrient due to its major biological roles: enzyme activation, maintenance of turgor, and regulation of internal pH [125]. Due to its important roles, there are many transport systems for transporting K^+ across membrane in response to osmotic stress [126]. In this study, the negative effect of high concentration of NH_4Cl (perhaps due to high osmolarity or medium ionic strength) is counteracted under high concentrations of NaH_2PO_4 and/or K_2SO_4. Gram-negative bacteria respond to high osmolarity by accumulating K^+ and other anions [127].

Furthermore, magnesium ion, the second most abundant cation is essential for bacterial growth and maintenance [128]. Its biological functions include: genomic stability, enzymes cofactor, and regulation of cell cycle [129]. Deficiency of Mg^{2+} is associated with disintegration of bacterial ribosomal subunits, which in turn causes membranes to be leaky [130]; this effect is more detrimental to Gram-positive bacterium. Although an essential element for bacterial growth, Mg^{2+} appears not to be limiting under laboratory condition—as observed in this study. This is due to unique transporters (constitutive ubiquitous CorA, first described in *Escherichia coli*, inducible less distributed MgtA, and MgtB extensively studied in *Salmonella typhimurium*), which scavenge and accumulate the micromolar concentration needed for

active growth [128,130]. Concentration as little as 4 μg/mL in a medium is sufficient to support the growth of most Gram-negative bacteria [131].

2.5 Conclusion

This study provides insight into the impact of media formulations on growth and cell density of *C. necator* H16. The growth model shows that the bacterium is able to grow to high-cell-density with minimal concentration of each growth component—provided the concentration of amino acids and trace elements are balanced. When cultivated under the media formulae generated in this study, growth peaked at 72 h and remained stationary after a further 20 h. High concentrations of fructose, amino acids and $CaCl_2$, and interactions between amino acids and fructose, fructose and $CaCl_2$, and amino acids and trace elements significantly impacted positive on cell density. While high concentrations of Na_2HPO_4 and trace elements, and interactions between fructose and Na_2HPO_4 were detrimental to growth. Addition of amino increased growth rate and resulted in reproducible high-cell-density across media. Besides fructose, only magnesium is considered essential for *C. necator* growth, whilst amino acids and K_2SO_4 are considered important. Thus, a defined medium supporting high cell density of *C. necator* lacking some components (i.e. Na_2HPO_4, $CaCl_2$, NH_4Cl, and trace elements) can still result in high-cell-density comparable to composition with all components. The statistical approaches undertaken in this study and the information gained will inform future efforts aimed at optimising other *C. necator* responses. This includes the biosynthesis of polyhydroxyalkanoates, platform chemicals, proteins and other high valued products, as well as future optimisation of growth under lithoautotrophic conditions using renewable resources.

Chapter 3.

Characterisation of Existing Broad Host Range Plasmid Parts for Bioengineering Applications in *Cupriavidus necator* H16

3.1 Introduction

Naturally occurring plasmids have contributed significantly to the field of molecular biology. They serve as means of transferring genetic information from one biological host to another, hence the name 'vectors' [132]. Due to the increasing number of microbes that have the potential of being deployed as microbial chassis for synthetic biology applications, it is important to characterised plasmid biological parts (bioparts) for compatibility and functions in a host of interest, prior to propagating any heterologous gene. A well-characterised biopart for a particular microbial host can be excised, modified and combined with other bioparts to generate plasmids that are more compatible and with improved function for another microbial host [133]. The use of well-characterised bioparts allows the prediction of a system output. The ability to predict the behaviour of individual biopart in a complex system together with rapid prototyping is vital to the success of synthetic biology application across disciplines [134].

While narrow host range plasmids are appealing in bioengineering applications owing to the small size (< 1 kb) of their replication origin (*ori*) and high copy number, they are unable to direct replication in non-model microbial chassis [Chapter 4]. Broad host range (BHR) plasmids are important tool for studying the genetics of diverse group of bacteria, as they are able to replicate across different bacterial groups. Unlike narrow host range plasmids, BHR plasmids have their own replication protein(s), which enables them replicate autonomously in different hosts thus achieving some level of species independence. Nevertheless, the expression level of a gene of interest (goi) carried on BHR plasmid is dependent on the compatibility of other bioparts with the expression host. Therefore, characterising biopart that are part of a BHR plasmid for a microbial chassis is key to understanding how plasmid propagation affects host metabolic and phenotypic features.

Two essential bioparts that require characterisation in *Cupriavidus necator* are replication origin (*ori*) and antibiotic resistant cassette (Ab^R). Few *ori* have been used to direct plasmid replication in *C. necator*. These *ori* (RSF1010, RP4/RK2, pMOL28 and pBBR1) have been instrumental in *C. necator* bioengineering applications [47–49,135]. Specifically, plasmids with pBBR1 *ori* are widely used perhaps due to their smaller size relative to other BHR *ori*.

However, pBBR1 *ori* is yet to be characterised for features fit for bioengineering applications in *C. necator*. Examples of such features are modularity and high-throughput assembly. These features are crucial in synthetic biology, as they enable exchange of bioparts in a modular format and efficient assembly of multiplex bioparts.

In addition, the range of antibiotic cassettes available for selecting *C. necator* transformants is limited. Kanamycin cassette is the most used for *C. necator* applications, with minimum effective concentration for selecting transformants not less than 200 *μ*g/mL[16,44,52]. Moreover, one of the major challenges with the existing BHR plasmids in *C. necator* is low transformation efficiency [56,73]. Existing BHR plasmids are built for conjugative transfer, which is less efficient compared to electroporation. Therefore, this study is aimed at characterising broad host range *ori* and antibiotic cassettes for applications in *C. necator*, to make these bioparts suitable for constructing modular plasmids that can be assembled in a high-throughput fashion and delivered to the bacterium by electroporation. Furthermore, this study is designed to test the hypothesis that low transformation efficiency of *C. necator* is due to incompatibility of bioparts of existing BHR plasmids with the bacterium.

3.2 Materials and Methods

3.2.1 Bacterial strains, media and antibiotics

High efficiency chemically competent *Escherichia coli* cloning strain, DH5α, was obtained from New England BioLabs (NEB UK). Carbenicillin, kanamycin and spectinomycin were obtained from Melford, ampicillin and erythromycin from Sigma-Aldrich, whilst chloramphenicol and tetracycline were obtained from Duchefa Biochemie. Gentamicin and lysogen broth (LB) powder were obtained from Formedium. Fructose was obtained from Sigma-Aldrich. Bacto tryptone and yeast extracts were obtained from BD Biosciences. Routine *E. coli* and *C. necator* cultivation (37°C and 30°C, respectively) was performed on LB medium (1% tryptone, 0.5 % yeast extract, 1% NaCl) under aerobic condition. When needed, LB medium (broth or agar) was supplement with appropriate concentrations of antibiotics (12.5 *μ*g/mL tetracycline; 25 *μ*g/mL chloramphenicol; 40 *μ*g/mL kanamycin) or (10 *μ*g/mL tetracycline; 50 *μ*g/mL chloramphenicol; 200 *μ*g/mL kanamycin) for *E. coli* and *C. necator*, respectively [51,59]. Super optimal broth (SOB) contained 2% Bacto tryptone, 0.5% Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$ and 10 mM $MgSO_4$. *E. coli* SOB with catabolite repression (SOC) contained 20 mM glucose, while that for *C. necator* was supplemented with 20 mM fructose in lieu of glucose.

3.2.2 Molecular cloning kits

Broad host range vectors, pKT230, maintained in *Escherichia coli* (K12 SK1592; DSM-5595) was obtained from DSMZ. Plasmid pBBR1MCS-2 was maintained in *E. coli* strain (K12 DH5α; NCCB 3234) obtained from Centraalbureau voor Schimmelcultures Netherlands collections. Lyophilised pBHR1 was obtained from MoBiTec GMBH, Germany. Primers were designed using Primer3 Plus and synthesised by Invitrogen, Life Technologies. Q5® High-Fidelity DNA polymerase (NEB UK) was used for all amplifications. Plasmid purifications were carried out using QIAGEN plasmid mini kit. Polymerase chain reactions were performed using a Bio-Rad thermal cycler. Purification of larger PCR products and restriction digests were carried out with QIAquick PCR purification kit, whilst smaller DNA fragments were purified using QIAquick nucleotide removal kit. Gel extraction and purification were carried out using QIAquick gel extraction kit. DNA quantification was carried out using a Qubit 2.0 Fluorometer (Invitrogen™, Life Technologies™). Molecular grade (DNase/RNase free) Just water (Clent Life Science) was used as the elution buffer. All constructs were verified by Sanger DNA sequencing. Primers used for the amplification of each biopart are as shown in Table 3.1.

Table 3. 1 Plasmids and bioparts.

Plasmid	Description	Source
pKT230	*nic oriV*; *Mob*; *SmR*; *SuR*	[135]
pBHR1	pBBR1 *Rep; Mob*; *KanR*; *CmR*	[49]
pBBR1MCS-2	pBBR1 *oriV-Rep*; *Mob*; *KanR*; MCS	[48]
pBBR1MSC-2a	pBBR1 *oriV-Rep*; *Mob*; *KanR*; MCS	This study
pBHR1ΔMob	pBBR1 *Rep; KanR*; *CmR*	This study
Primers		
pBHR1_KanR	F: 5'-**T^CCCGGA**GTCCCGTCAAGTCAGCGTAA-3' R: 5'-**A^GATCT**GGAAAGCCACGTTGTGTCTC-3'	
pBHR1ΔMob	F: 5'-**A^GATCT**CGAGCTTGACCACAGGGATT-3' R: 5'-**A^GATCT**GCCCTTTGCGCCGAATAAAT-3'	

pBBR1MCS-2a is a hybrid of pBBR1MCS-2 and pBHR1. pBHR1ΔMob was derived by removing *Mob* from pBHR1.

3.2.3 Electroporation

Electrocompetent cells of *C. necator* was prepared as described [74,136]. Briefly, a 48 h colony was inoculated into 10 mL SOB and incubated overnight at 30°C, 200 rpm. Two hundred and fifty microliter of the overnight culture each was transferred into fresh 250 mL SOB contained in 1 L flasks. The inoculated flasks were incubated accordingly, and growth monitored until cultures reached the desired OD_{600nm}, 0.4–0.8. Cells were harvested at 3000 x g for 10 min using an ultra-centrifuge (RC5C Sorvall). The harvested cells were washed twice with 250 mL 10% ice-cold glycerol at 3000 x g, 10 min, 4°C. Electrocompetent cells were pooled and

concentrated in residual buffer. Aliquots of 100 µL were transferred into 0.5 mL ice-cold microfuge tubes and immediately stored at -80°C.

To transform *C. necator,* 50 µL electrocompetent cells were mixed with ~50 ng of purified plasmid. The mixture was transferred into 2 mm gap ice-cold cuvette and pulsed using a Bio-Rad GenePulser Xcell™. The electroporation conditions were: 2.5 kV, 25 µF, 200 Ω, and a single pulse as described [20]. Immediately, 450 µL SOC (SOB supplemented with 20 mM fructose) was added into the cuvette containing the pulsed cells. The suspension was subsequently transferred into 2 mL microfuge tube and incubated for 2 h at 30°C, 200 rpm. Following the incubation, 100 µL of culture was diluted and spread on LB agar supplemented with appropriate antibiotics and incubated at 30°C for 48 h. Transformation efficiencies (T.E., transformants/µg DNA) were determined as:

$$\text{Transformation efficiency (transformants/}\mu\text{g DNA)} = \left[\text{Transformants} \times \frac{\text{Total volume of transformation reaction } (\mu\text{L})}{\text{Volume spread on agar plate } (\mu\text{L})} \times \text{Further Dilution factor (if any) before spreading on agar plate}\right] \times \frac{1}{\text{Amount of DNA delivered } (\mu\text{g})}$$

The variables in square bracket is equivalent to colony forming unit (CFU). Thus, T.E. can also be expressed as CFU/µg DNA

3.2.4 Heat-shock transformation

C. necator was grown and harvested as described previously for electroporation. However, harvested cells were suspended in buffers containing different cations [137]. In the first wash, cells were incubated for 20 min on ice in 250 mL appropriate ice-cold wash buffer. Following this, the cell suspension was concentrated by centrifuging at 2000 x g for 10 min. Next, the cell pellet was suspended in 250 mL ice-cold wash buffer (containing 10% glycerol and the same cation concentration as the first wash) and centrifuged for 2000 x g for 10 min. Subsequently, the bacterial pellet was suspended in residual buffer and combined. Aliquots of 100 µL were transferred into 2 mL ice-cold microfuge tubes, flash frozen in liquid nitrogen and stored at -80°C. For heat-shock transformation, 50 µL of chemically competent *C. necator* cells were mixed with 50–1000 ng plasmid DNA and incubated on ice for 30 min. Thereafter, heat-shock was performed at different temperatures for 90 s, and 200 µL SOC was added afterwards. Recovery was carried out as described previously for electroporation. After the recovery, 100 µL of undiluted cells were plated on LB agar supplemented with the appropriate antibiotics and incubated at 30°C for 48 h. *C. necator* transformants were

confirmed by a colony PCR and plasmid bioparts confirmed by sequencing. Compositions of buffer for preparation of chemically competent cells are given in Table 3.2.

Table 3. 2 Buffer for preparation of chemically competent *C. necator*.

Buffer	Chemical composition
Control (A)	100 mM $CaCl_2$, 100 mM $MgCl_2$, 85 mM $CaCl_2$ and 15% glycerol
B	30 mM $CaCl_2$
C	15 mM $CaCl_2$, 250 mM KCl, 55 mM $MnCl_2.4H_2O$, 0.5 M PIPES
D	15 mM $CaCl_2$, 250 mM KCl, 55 mM $MnCl_2.4H_2O$, 10 mM Hepes
E	Buffer I: 50 mM $CaCl_2$, 50 mM $MnCl_2$, 30 mM KOAc; buffer II: 75 $CaCl_2$, 10 mM MOPs, 15% glycerol
F	1 M $MgCl_2$, 5 g PEG 8000, 2.5 mL DMSO, Lysogeny broth

References to details for buffer preparations are provided in Appendix B.

3.2.5 Substitution of kanamycin cassette (KanR) in pBBR1MCS-2 with KanR from pBHR1

The *KanR* in pBHR1 was amplified with a forward primer containing PfoI restriction sequence (F: 5'-**T^CCCGGA**GTCCCGTCAAGTCAGCGTAA-3') and a reverse primer containing a Bg1II restriction sequence (R: 5'-**A^GATCT**GGAAAGCCACGTTGTGTCTC-3'). Following amplification, the insert (pBHR1 *KanR* amplicon) was digested with Bg1II and PfoI FastDigests (Thermo Scientific™) and purified using a QIAquick PCR purification kit. Backbone vector, pBBR1MCS-2, was treated with the same restriction enzymes and gel purified. Ligation of the insert and backbone vector was performed at 1:1 ratio using a T4 DNA ligase (Thermo Scientific™), and the ligation reaction incubated overnight at room temperature. The resulting plasmid designated pBBR1MCS-2a was transformed into *E. coli*, purified and 50 ng delivered to *C. necator* by electroporation.

3.2.6 Removal of mobilisation sequence (Mob) from pBHR1 by PCR cloning

Although not visible when viewed using a SnapGene tool, the *Mob* sequence in pBHR1 spans 1–799 bp and 5080–5300 bp (i.e. 5300–799 region) [49], whilst that in pBBR1MCS-2 spans 1–799 bp and 4945–5144 bp (i.e. 5144–799 region) [48]. The forward (F: 5'-**AGATCT**CGAGCTTGACCACAGGGATT-3') and the reverse primer (R: 5'-**AGATCT**GCCCTTTGCGCCGAATAAAT-3') primers have Bg1II recognition sequence to allow self-hybridisation of the amplicons. Both were designed to amplify from the region closest to the target and moving downstream, away from the target, in the opposite directions. Experimental procedures: amplification, gel extraction and purification, restriction digests with Bg1II, ligation, transformation into *E. coli*, plasmid purification and transformation into *C. necator* were carried out as described previously.

3.2.7 Antimicrobial susceptibility test

Minimum inhibitory concentrations (MIC) of different antibiotics were determined for *C. necator* as described by the European Committee for Antimicrobial Susceptibility Testing [138]. Briefly, 100 µL Mueller-Hinton (M-H) broth (0.2 % beef extract, 1.75 % casein hydrolysate and 0.15% starch) was transferred into 96-well microtiter plate (Fig. 3.1). The designated 'start' well (column 2) of each antibiotic contained 100 µL double-strength (2X) broth, whilst the remainder (column 3–11) contained 100 µL single-strength (1X) broth (Fig. 3.1A). Next, 100 µL (2000 µg/mL) each antibiotic (ampicillin, carbenicillin, chloramphenicol, erythromycin, gentamicin, kanamycin, streptomycin and tetracycline) dissolved in appropriate solvent was aseptically transferred into the corresponding 'start' well (Fig. 3.1B). From the 'start' well, two-fold serial dilution was performed through to the designated 'last' well (Fig. 3.1B). This was achieved by transferring 100 µL from the 'start' well—now containing a total of 200 µL solution made of up 1X M-H broth and 1000 µg/mL antibiotic—into the next well, and from this well, 100 µL was transferred to the subsequent well (Fig. 3.1B). After the serial dilution, 100 µL was removed from the designated last well containing 1.95 µg/mL antibiotic to adjust the volume to 100 µL. Subsequently, 100 µL of adjusted inoculum (1.0×10^6 cfu/mL, in M-H broth) was transferred into each well, bringing the total volume to 200 µL (Fig. 3.1C). This resulted in a further two-fold serial dilution of each antibiotic concentration. Hence, the final concentration of each antibiotic in the wells were: 500, 250, 125, 62.5, 31.25, 15.63, 7.81, 3.91, 1.95, 0.98 µg/mL, and the final inoculum concentration was 5.0×10^5 cfu/mL.

Two control wells, one containing 100 µL 1X M-H broth and 100 µL adjusted inoculum (columns 1) and another containing 200 µL 1X M-H broth (columns 12), were used as standards (Fig. 3.1C and D). Serial dilution of antibiotics and inoculum transfer were carried out using an automated liquid handling system (Eppendorf epMotion M5073). The MIC for each antibiotic was determined as the lowest concentration that inhibited growth after 20 h incubation at 30°C, 180 rpm. Optical density (OD_{600nm}) was measured using a Varioskan LUX™ Multimode Microplate reader (Thermo Scientific).

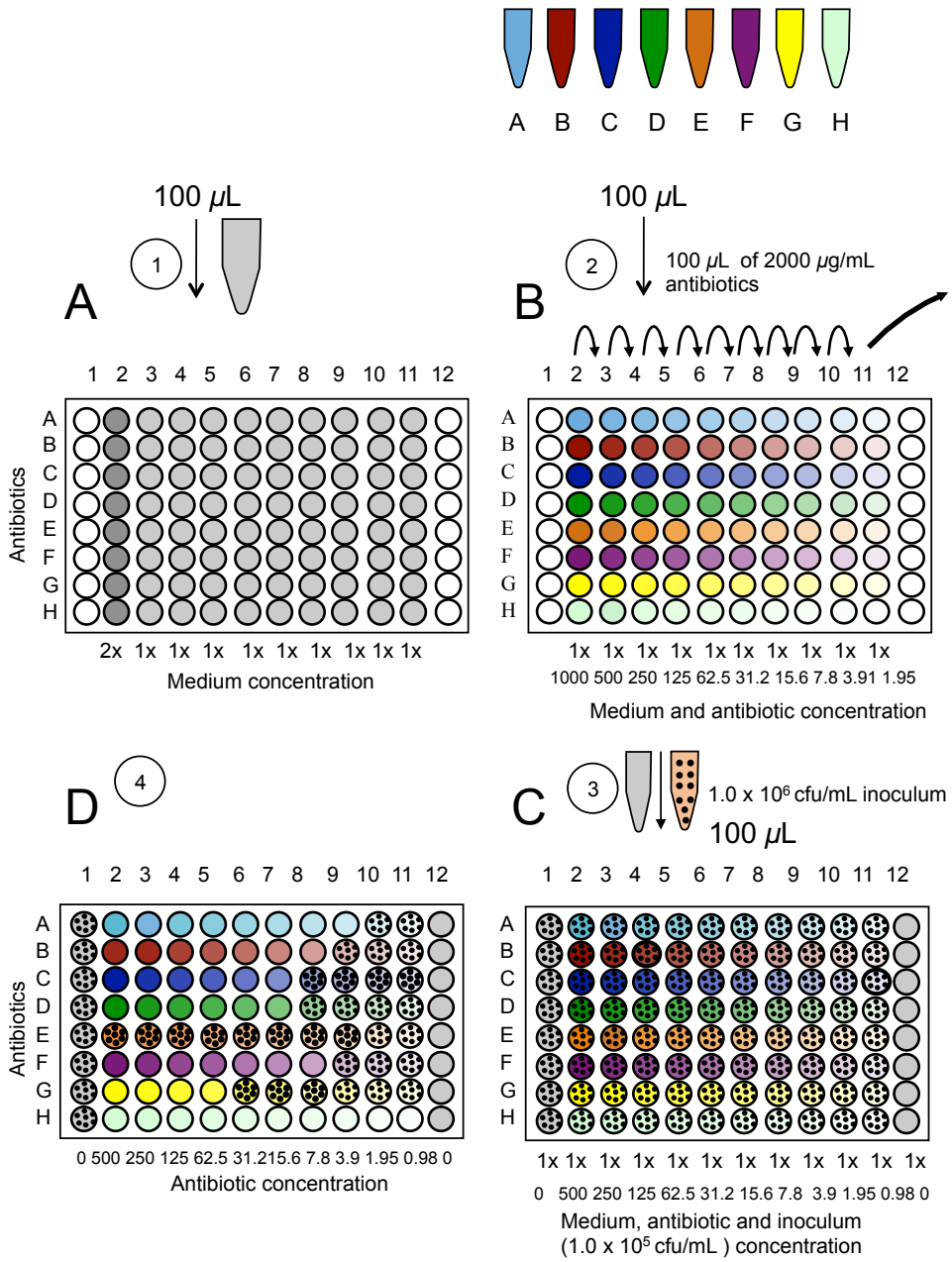


Fig. 3. 1 Determination of minimum inhibitory concentration (MIC).

1. Transfer of 100 µL Muller-Hinton (M-H) broth to 96-well microtiter plates; dark shaded grey wells contained double strength (2x) broth; light grey wells contained single strength (1x) broth; white well contained no broth at this stage. **2**. Transfer of 100 µL 2000 µg/mL antibiotics solution into 'start' wells (column 2), and subsequent serial dilution across wells (columns 3–11), 100 µL (curved arrow) was discarded from column 11 for each row to adjust the volume to 100 µL. **3**. Transfer of 100 µL single strength (1x) broth into control wells (columns 1 and 12), followed by transfer of 100 µL adjusted inoculum into wells, columns 1–11; column 12, which no inoculum was added served as negative control, whilst column 1 to which inoculum was added but contained no antibiotic served as the positive control. **4.** Examination of plate at 20 h incubation at 30°C. Plates: **A**. plate containing 2x and 1x broth; **B**. Plate containing serially diluted antibiotic solution; **C**. Plate containing antibiotic solution and inoculum, further diluted by two-fold; **D**. MIC of each antibiotic on *C. necator*. Antibiotics: Ampicillin (A), Carbenicillin (B), Chloramphenicol (C), Erythromycin (D), Gentamicin (E), Kanamycin (F), Spectinomycin (G) and Tetracycline (H). The closed dark circles represent microbial growth. Same procedure was carried out for both broth and colony suspension method of MIC, *n* = 3 biological replicates.

To determine the mode of action of each antibiotic (MoA) by minimum bactericidal concentration (MBC), 100 μL from wells that showed no evidence of growth (i.e. from the MIC well upstream) were spread on M-H antibiotic-free agar and incubated at 30°C, 48 h (Fig. 3.2). Lastly, guided by the MIC results, dose response effects were determined for each antibiotic within 0–200 μg/mL concentration range, at 1 mL cultivation scale, using 48-well plates. The same inoculum concentration (1.0 x 10^6 cfu/mL) was spread on freshly prepared M-H agar supplemented with appropriate antibiotics to determine the colony-forming unit. Unlike the MIC, microtiter well plates were incubated for 24 h, whilst agar plates were incubated for 48 h, both at 30°C.

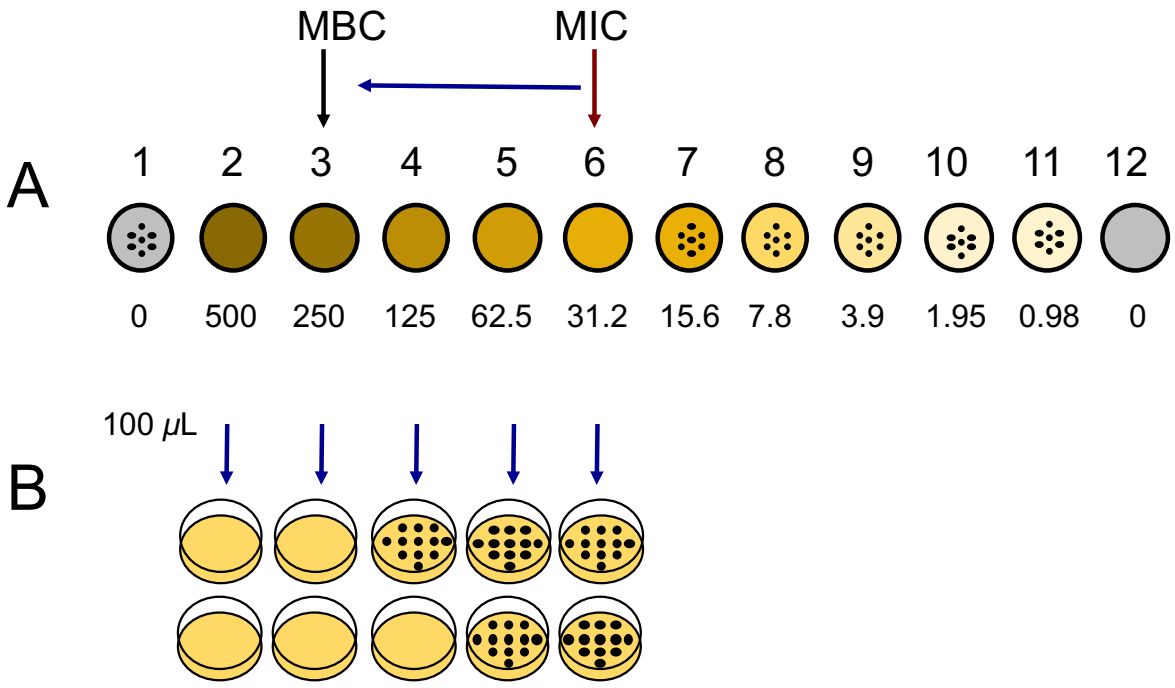


Fig. 3. 2 Minimum bactericidal concentration (MBC).

A. Microtiter well plate (column) following minimum inhibitory concentration (MIC). **B.** Transfer of 100 μL solution from wells that showed no evidence of growth into antibiotic-free Muller-Hinton (M-H) agar plates. Plates were examined after 48 h. The minimum concentration of well that showed no growth on agar plate after incubation is designated the minimum bactericidal concentration. Antibiotic that allowed growth on M-H agar plates at high concentration is described as biostatic, whilst those that allowed no growth is described as having bactericidal effect. The closed dark dots represent microbial growth.

3.3 Results

3.3.1 Identification of broad host range plasmids suitable for transforming C. necator

C. necator was first screened for its ability to be transformed by electroporation using existing broad host range (BHR) plasmids: pBHR1, pBBR1MCS-2 and pKT230. These plasmids carry kanamycin resistant cassette (*KanR*) as the selection marker (Fig. 3.3 A-C). Only pBHR1 transformed *C. necator* when selected at 200 μg/mL kanamycin. However, when selected at 100 μg/mL kanamycin, a small number of colonies were recovered for *C. necator* transformed with each of the plasmid (Fig. 3D). The colonies were confirmed to be transformants by a colony PCR. Nevertheless, with pBHR1 the transformation efficiency obtained at 200 μg/mL kanamycin was considerably higher than that obtained at 100 μg/mL kanamycin (Fig. 3D).

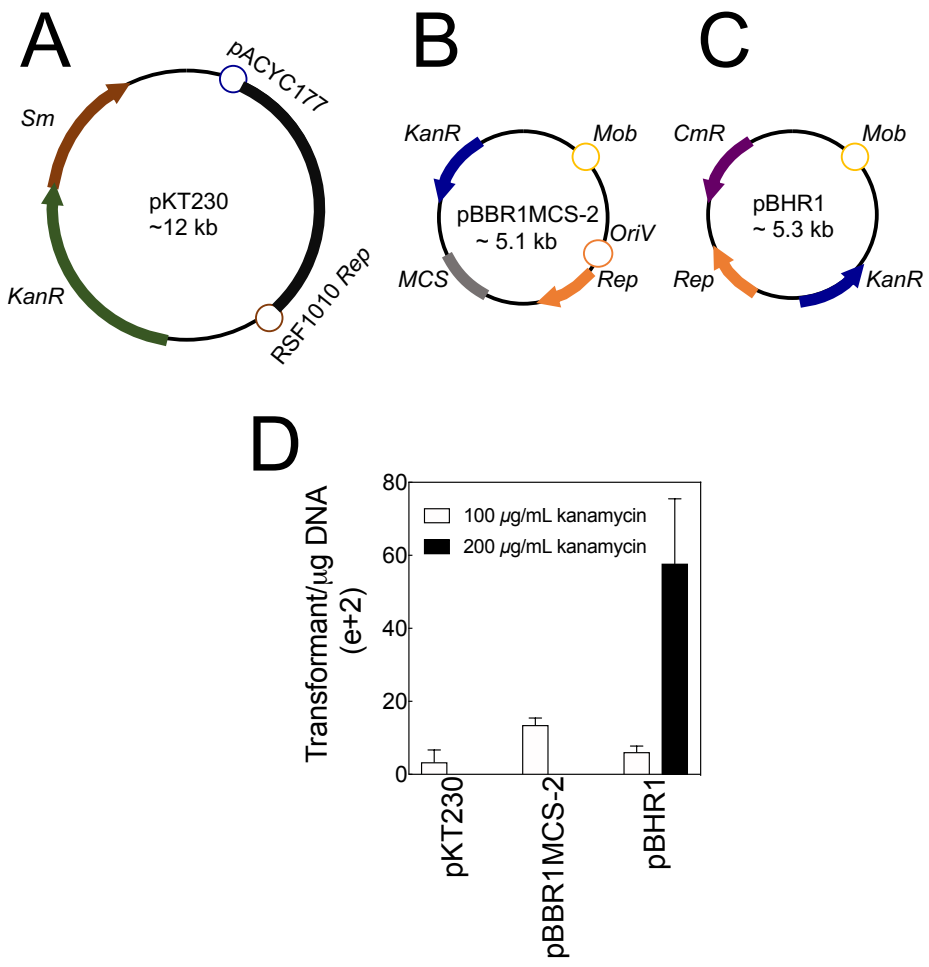


Fig. 3. 3 Transformation of *C. necator* with existing broad host range plasmids.

A. pKT230 with large replication sequence. **B.** pBBR1MCS-2 with multiple cloning site (MCS) for insertion of goi. **C.** pBHR1 derived by removing a frame shift in pBBR122 [49]. **D.** Transformation efficiencies of *C. necator* with the existing BHR plasmids, with transformants selected at 100 or 200 μg/mL kanamycin. Error bars are S.E.M., *n* = 3 biological replicates.

Given that the plasmids each carry *KanR*, along with pBHR1 and pBBR1MCS-2 seemingly having the same replication machinery (pBBR1 Rep), the ineffective transformation of the bacterium with either pBBR1MCS-2 or pKT230 could not be attributed to either the replication sequence or antibiotic selection. Nonetheless, pKT230 is more than twice the size of pBHR1 and pBBR1MCS-2. Due to the large size of pKT230, attention was shifted to pBHR1 and pBBR1MCS-2 to achieve the objective of building modular minimal plasmids.

3.3.2 Heat-shock transformation of Cupriavidus necator

With the understanding that no study has reported successful transformation of *C. necator* by heat-shock transformation, attempts were made to transform the bacterium by this method using pBHR1 and pBBR1MCS-2. Intriguingly, both plasmids transformed the bacterium and colonies on 200 *μ*g/mL kanamycin-LB agar plates were confirmed to be transformants by a colony PCR. However, the transformation efficiencies were very low compared to that obtained by electroporation (Fig. 3.4A). Because pBHR1 yielded more transformants, further demonstration of heat-shock transformation—with additional new set of buffers for preparing *C. necator* chemically competent cells, and cells heat-shocked at different temperatures (40–60°C)—were carried out with this plasmid. Chemically competent *C. necator* cells prepared in three of the five additional transformation buffers were successfully transformed by heat-shock (Fig. 3.4B). This resulted in a total of four (the initial and three additional) buffers that can potentially be used for preparing *C. necator* chemically competent cells. The bacterium is susceptible to heat-shock between 45–50°C, with > 90 s heat-shock interval. However, the major limitation of this method of transformation is the amount of DNA (approximately 1 *μ*g) required to transform the cells in comparison to that required (~ 50 ng) for obtaining higher transformation efficiency by electroporation. Both pBHR1 and pBBR1MCS-2 are low or medium copy plasmids and attempts to obtain high plasmids yield after purification, without further concentration, was unsuccessful. Therefore, electroporation was chosen as the most efficient means of delivering plasmids to *C. necator*.

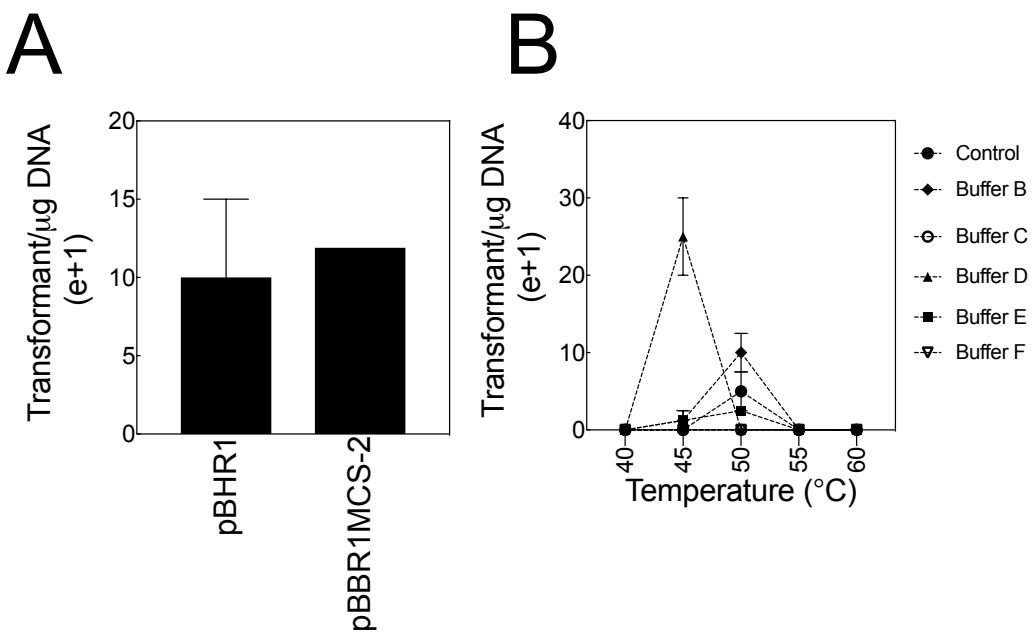


Fig. 3. 4 Demonstration of heat-shock transformation of *C. necator* with existing broad host range plasmids.

A. Transformation efficiency of *C. necator* with pBHR1 and pBBR1MCS-2 following heat-shock transformation. Chemically competent cells were first washed in 100 mM $MgCl_2$, followed by a second wash in 100 mM $CaCl_2$ and a final wash in 85 mM $CaCl_2$ with 15 % glycerol v/v. This buffer subsequently served as control. **B.** Further demonstration of heat-shock transformation of *C. necator* using different sets of chemically competent cells prepared in buffers with different chemical compositions, and heat-shocked at different temperatures. *C. necator* transformants were selected on 200 *μ*g/mL kanamycin-LB agar plates. Bars are S.E.M., *n* = 2 biological replicates. References to compositions of each buffer are given in the Appendix B.

3.3.3 Kanamycin resistant cassette in pBBRMCS-2 is responsible for ineffective transformation of C. necator by electroporation

Due to the observed differences in the transformation efficiency of *C. necator* with pBHR1 and pBBR1MCS-2 by electroporation, these two plasmids were further investigated. Structurally, both plasmids have similar bioparts: *KanR* and *Mob*. The notable difference is the presence of chloramphenicol resistant marker (*CmR*) in pBHR1 and multiple cloning sites (MCS) in pBBR1MCS-2 (Fig. 3.3 B and C). Although both plasmids were originally mapped as having the same replication sequence (*Rep*) directing replication, SnapGene shows both plasmids to differ significantly in the replication machinery (Fig. 3.5 A and B). It appears pBBR1MCS-2 has both *Rep* and *oriV* sequences, whilst pBHR1 has only *Rep* sequence. The pBBR1 *oriV* is the replication sequence from *Bordetella bronchiseptica*, which requires the replication protein (Rep) to direct autonomous replication in a host, hence broad host range replicon.

More importantly, the kanamycin DNA and protein sequences in both plasmids are entirely different (Table 3.3). The *KanR* in pBHR1 codes for aminoglycoside *O*-phosphotransferase class I, subtype 'a' (*APH (3')-Ia*), whilst that in pBBR1MCS-2 codes for aminoglycoside *O*-phosphotransferase class II, subtype 'a' (*APH (3')-IIa*). Therefore, it was hypothesised that the ineffective transformation of *C. necator* with pBBR1MCS-2 by electroporation is due to the *KanR*. Further, considering that *C. necator* pBBR1MCS-2 transformants were recovered when selected at 100 *μ*g/mL on LB-agar plates, it was proposed that the difference in enzyme specificity is responsible for effective transformation of *C. necator* with pBHR1 and ineffective transformation with pBBR1MCS-2 under electroporation conditions. It is plausible that the bacterium took up pBBR1MCS-2 after electroporation but failed to code sufficient resistance to kanamycin within the short period of recovery, 2 h. This subsequently prevented transformants forming at high concentration of kanamycin (200 *μ*g/mL).

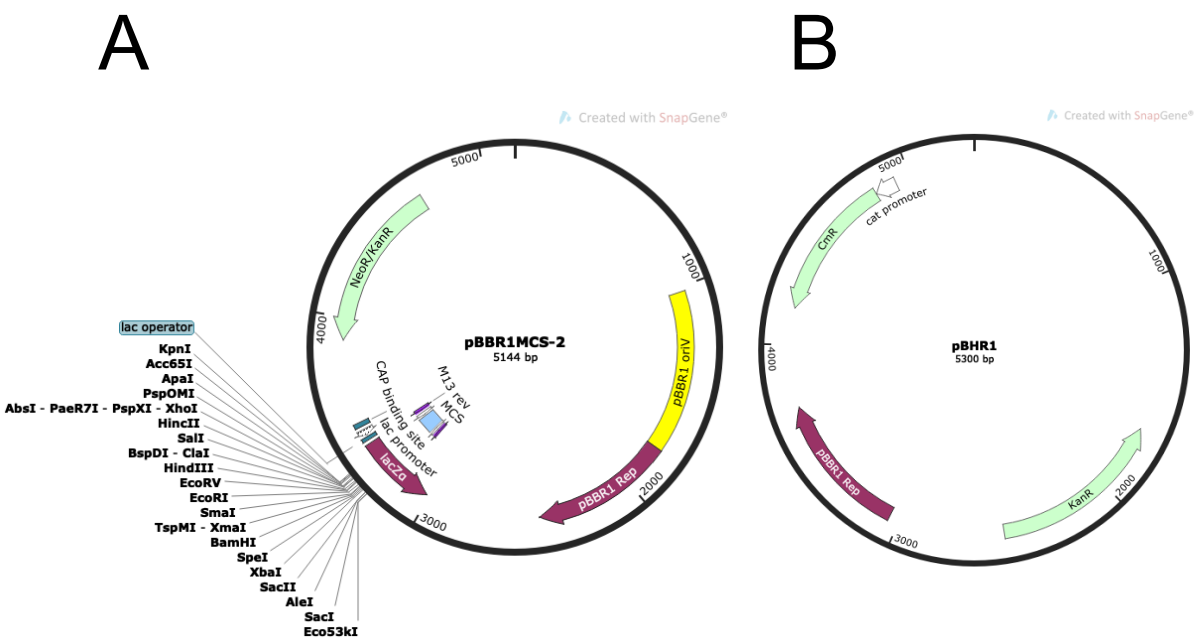


Fig. 3. 5 SnapGene map of widely used broad host range plasmids for *C. necator* application.

A. pBBR1MCS-2 depicting four bioparts. **B.** pBHR1 depicting three bioparts. It is noteworthy that both plasmids have mobilisation sequence (*Mob*), essential for conjugative transfer. This region spans 5080–799 bp and 5144–799 bp in pBHR1 and pBBR1MCS-2, respectively.

Table 3. 3 Alignment of kanamycin resistant cassettes.

```
pBHR1_KanR           MSHIQRETSCSRPRLNSNMDADLYGYKWARDNVGQSGATIYRLYGKPDAPELFLKHGKGS     60
pBBR1MCS-2_KanR      --MI--EQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQ-GRPVLFVKTDLSG     55
                        *  * .  :    :     *:**.**::.:* *.*:::** .: . * **:* . ..

pBHR1_KanR           VANDVTDEMVRLNWLTEF-MPLPTIKHFIRTPDDAWLLTTAIPGKTAFQVLEEYPDSGEN     119
pBBR1MCS-2_KanR      ALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDLLSSHLA----PAE     111
                     . *:: ** .**.**:   :*  :: ..:   .  ***   :**:  :.          :

pBHR1_KanR           IVDALAVFLRRLHSIPVCNCPFNSDRVFRLAQAQSRMNNGLVDASDFDDERNGWPVEQVW     179
pBBR1MCS-2_KanR      KVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDEEHQGLAPAELF     171
                      *. :*  :****::  ..***: :  .*: :*::**: **** .*:*:*::*    :::

pBHR1_KanR           KEMHKLLPFSPDSVVTHGDFSLDNLIFDEGKLIGCIDVGRVGIADRYQDLAILWNCL-GE     238
pBBR1MCS-2_KanR      ARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRYQDIALATRDIAEE     231
                     .::  :* . * ****** .* *::.::*:: * ** **:*:******:*:  . :  *

pBHR1_KanR           FSPSLQKRLFQKYGIDNPDMNKLQFHLMLDEFF   271
pBBR1MCS-2_KanR      LGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF   264
                     :. .  .*::  ***  ** ::: *: :*****
```

pBHR1_KanR codes for (*APH (3')-Ia*), while pBBR1MCS-2_KanR codes for (*APH (3')-IIa*). Plasmid accession: pBHR1, Y14439.1; pBBR1MCS-2, U23751. Protein sequences were aligned using CLUSTAL O (1.2.4).

Aware that pBHR1 and pBBR1MCS-2 differ in their replication machinery and in other bioparts, the hypothesis that *KanR* in pBBR1MCS-2 is responsible for ineffective transformation of *C. necator* with this plasmid was tested. This hypothesis was tested adopting conventional cloning method to substitute *KanR* in pBBR1MCS-2 with that from pBHR1. Remarkably, the resulting hybrid plasmid (pBBR1MCS-2a) transformed *C. necator* with high transformation efficiency. Moreover, the transformation efficiency obtained for pBBR1MCS-2a was two-fold higher in comparison to that obtained for pBHR1 (Fig. 3.6). The resulting transformants were confirmed by a colony PCR, and pBHR1 *KanR* in pBBR1MCS-2a confirmed by DNA sequencing. Hence, it is established that the ineffective transformation of *C. necator* with pBBR1MCS-2 is due to the *KanR* in the plasmid and not as a result of other differences in bioparts in comparison to pBHR1.

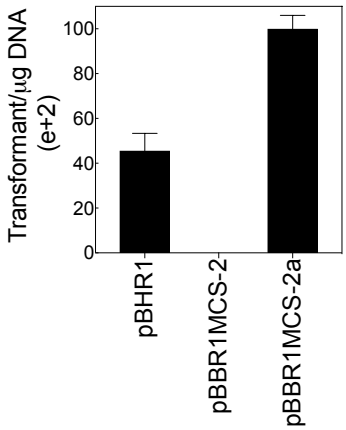


Fig. 3. 6 Difference in kanamycin resistant cassette on *C. necator* transformation.

pBBR1MCS-2a is derived by substituting *KanR* in pBBR1MCS-2 with *KanR* from pBHR1.

3.3.4 Mobilisation sequence in pBHR1 is dispensable for transformation of C. necator by electroporation

Guided by the objective of constructing modular minimal plasmids that can be delivered to *C. necator* by electroporation, the effect of excluding *Mob* sequence on plasmid transformation efficiency was examined. pBHR1 lacking *Mob* (ΔmobpBHR1) transformed *C. necator* with efficiency comparable to that of pBHR1 with *Mob* (Fig. 3.7). Therefore, it is established that *Mob* in pBHR1 is not essential for efficient transformation of *C. necator* by electroporation.

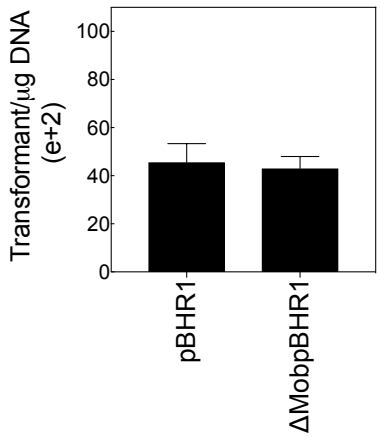


Fig. 3. 7 Removal of mobilisation (*Mob*) sequence did not affect *C. necator* transformation efficiency by electroporation.

3.3.5 Expanding antibiotic cassette for Cupriavidus necator

Having demonstrated that variation in the sequence of *KanR* can impact on the transformation efficiency, and that *Mob* sequence is not essential for transformation by electroporation, the susceptibility of *C. necator* to other antibiotics was determined. *C. necator* was tested against ampicillin, carbenicillin, chloramphenicol, erythromycin, gentamicin, kanamycin, spectinomycin and tetracycline. The presumptive antimicrobial susceptibility test, MIC, showed the bacterium to be susceptible to ampicillin, carbenicillin, chloramphenicol, erythromycin, kanamycin and tetracycline both for broth and colony suspension method of MIC determination. The bacterium displayed high degree of susceptibility to tetracycline, with MIC < 0.98 *μ*g/mL (Table 3.4). Gentamicin and spectinomycin showed weak antimicrobial activities against the bacterium even at a higher concentration (500 *μ*g/mL), indicating that the bacterium is resistant to both antibiotics. To determine how each antibiotic affected the bacterium, a minimum bactericidal concentration (MBC) assay was performed (Fig. 3.2). Ampicillin and carbenicillin appear to exert bactericidal effect, whilst chloramphenicol, erythromycin, kanamycin and tetracycline were

bacteriostatic in inhibiting growth. As expected, colonies were recovered on gentamicin and spectinomycin agar plates further indicating the resistance of *C. necator* to these antibiotics.

Table 3. 4 Determination of minimum inhibitory concentration of *C. necator* to antibiotics

Strain and antibiotics	MIC (*μ*g/mL) for broth culture method			MIC (*μ*g/mL) for colony suspension method		
C. necator H16	Run 1	Run 2	Run 3	Run 1	Run 2	Run 3
Ampicillin	1.95	3.91	1.95	< 0.98	3.9	1.95
Carbenicillin	7.81	7.81	3.91	3.91	7.81	3.91
Chloramphenicol	15.63	15.63	15.63	7.81	7.81	15.63
Erythromycin	15.63	31.25	7.81	31.25	15.63	31.25
Gentamicin	> 500	> 500	> 500	> 500	> 500	> 500
Kanamycin	7.81	7.81	7.81	7.81	15.63	7.81
Spectinomycin	250	> 500	62.50	> 500	> 500	125
Tetracycline	< 0.98	< 0.98	< 0.98	< 0.98	< 0.98	< 0.98

To confirm the results of the presumptive test (MIC), dose response assay to quantify the effect of each antibiotic on inhibiting *C. necator* growth was carried out at higher culture volume, 1 mL. Each antibiotic was tested at: 0, 50, 100 and 200 *μ*g/mL guided by the results of the MIC. Tetracycline, however, was tested at lower concentrations: 0, 0.5, 5, 20 and 50 *μ*g/mL. The susceptibility pattern observed at this confirmatory phase was similar to that observed during the presumptive phase. There was gradual decrease in growth with increasing concentration of antibiotics expect for tetracycline, which inhibited the same level of growth regardless the concentration (Fig. 3.8). Unlike the MIC, the bacterium displayed inconsistent susceptibility to ampicillin at 1 mL, growing at 200 *μ*g/mL of ampicillin. Chloramphenicol and erythromycin effectively inhibited growth at 50 *μ*g/mL, whilst carbenicillin and kanamycin inhibited growth at 100 *μ*g/mL (Fig. 3.8).

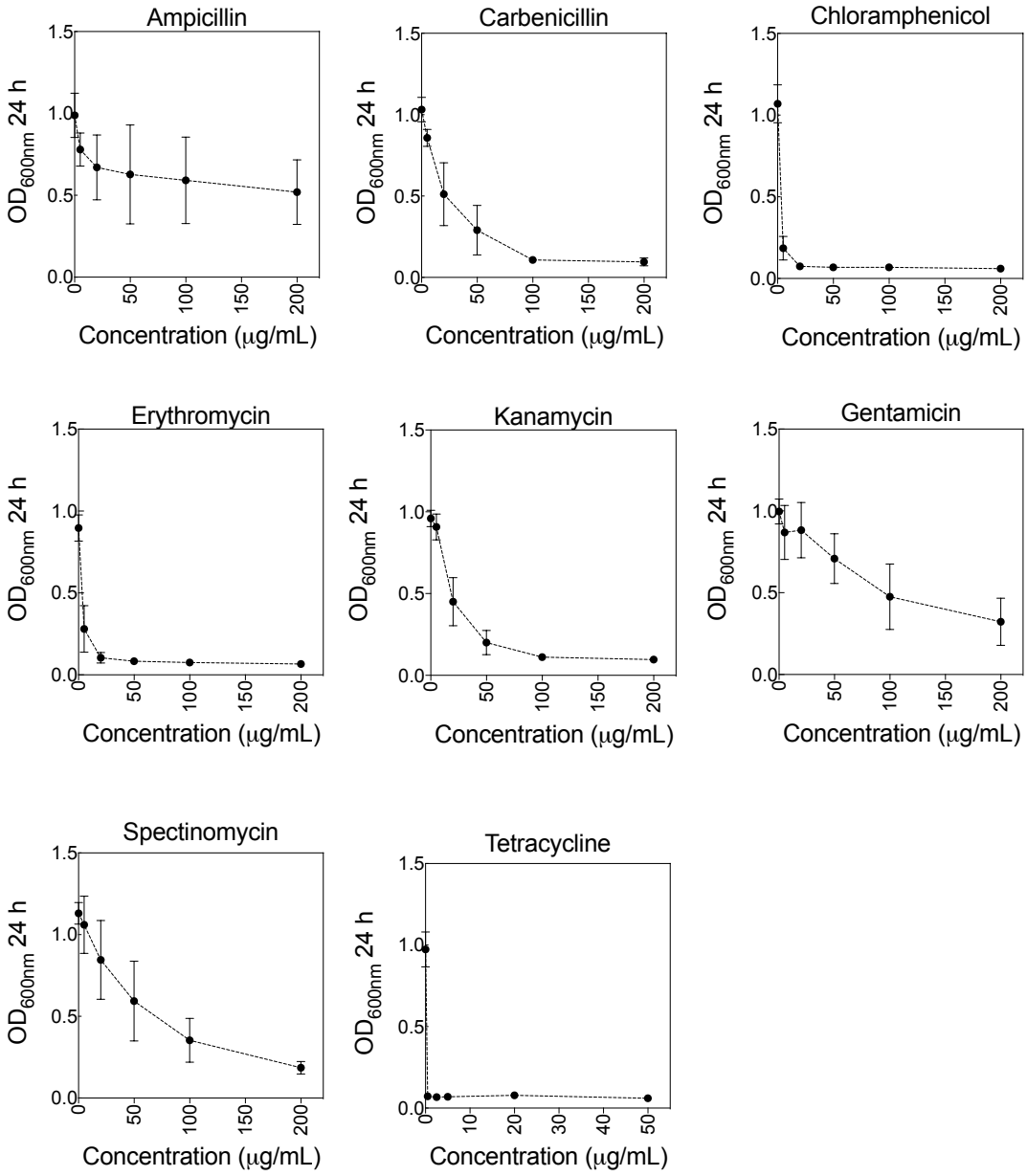


Fig. 3. 8 Determination of effective concentration for selecting *C. necator* transformants in broth.

Concentration tested were 0–200 *μ*g/mL except for tetracycline, which was tested between 0–50 *μ*g/mL. Error bars are S.E.M., *n* = 3 biological replicates.

Lastly, to complete the antimicrobial susceptibility, adjusted inoculum (1.0 x 10^6 cfu/mL) was spread on M-H agar supplemented with the same antibiotic concentration as described for the confirmatory test. As expected, the bacterium displayed considerable degree of resistance to gentamicin; colonies formed on M-H agar supplemented with 200 *μ*g/mL gentamicin (Table 3.5). In general, *C. necator* grew on M-H agar plates supplemented with 5

μg/mL of each antibiotic except for tetracycline plates, which inhibited colony formation even at concentration as low as 0.5 μg/mL (Table 3.5).

Following our systemic antimicrobial susceptibility screening (presumptive, confirmatory and completed phase), it was deduced that spectinomycin and gentamicin are unsuitable for selecting *C. necator* transformants. Ampicillin although very stable during the MIC, proved unstable at larger culture volume by yielding inconsistent results at the same antibiotic concentration. Carbenicillin, chloramphenicol, erythromycin, kanamycin and tetracycline are considered suitable for selecting *C. necator* both in liquid culture and on agar plates (Table 3.6).

Table 3. 5 Determination of effective concentration for selecting *C. necator* transformants on agar plates.

Strain or antibiotics	CFU/mL at different antibiotic concentrations			
C. necator H16	5 μg/mL	50 μg/mL	100 μg/mL	200 μg/mL
Ampicillin	$9.0 \pm 0.8 \times 10^5$	1.0 x10	0	0
Carbenicillin	$2.7 \pm 0.3 \times 10^6$	0	0	0
Chloramphenicol	2.0 x10	0	0	0
Erythromycin	6.0 x10	0	0	0
Gentamicin	$4.8 \pm 0.7 \times 10^6$	$2.67 \pm 0.05 \times 10^6$	$1.4 \pm 1.3 \times 10^3$	1.0 x10
Kanamycin	$5.3 \pm 1.0 \times 10^6$	3.0 ± 0.0 x10	0	0
Spectinomycin	$6.2 \pm 1.0 \times 10^6$	$2.3 \pm 1.3 \times 10^2$	0	0

The values shown are the mean of *n* = 3 biological replicates; error is S.E.M. The inoculum density for the control plate containing no antibiotics was $6.5 \pm 0.5 \times 10^6$ cfu/mL. It is noteworthy that tetracycline was tested at: 0.5, 5, 20 and 50 μg/mL and no *C. necator* colony was recovered after *n* = 3 biological replicates.

Table 3. 6 Summary of antimicrobial susceptibility of *C. necator*.

Antibiotic	MIC (μg/mL)	MoA (μg/mL)	$E_{conc.}$ (μg/mL)	Inference
Ampicillin	1.9 – 3.91	Bactericidal	> 100	Marginally suitable
Carbenicillin	3.91 – 7.81	Bactericidal	> 100	Suitable
Chloramphenicol	7.81 – 15.63	Bacteriostatic	50	Suitable
Erythromycin	7.81 – 31.25	Bacteriostatic	50	Suitable
Gentamicin	> 500	Resistant	Resistant	Not suitable
Kanamycin	7.81 – 15.63	Bacteriostatic	200	Suitable
Spectinomycin	> 62.50	Resistant	Resistant	Not suitable
Tetracycline	< 0.98	Bacteriostatic	< 10	Suitable

Inferences were drawn following systematic approach to three major phases of antimicrobial susceptibility tests: presumptive (MIC), confirmatory (MBC) and completed (culturing on agar plate supplement with antibiotic). Effective concentration for selecting *C. necator* transformants in broth or on agar is designated E_{conc}.

3.4 Discussion

Transformation of *Cupriavidus* with broad host range plasmids is mostly achieved via conjugation [56]. This approach relies on a second donor bacterium, *E. coli* S17-1. Among the three broad host range plasmids tested in this study, only pBHR1 effectively transformed *C. necator* by electroporation. Although plasmids with pBBR1MCS-2 and pKT230 backbones were reported to transform *C. necator* via electroporation [56,74,110], unsuccessful transformation of the bacterium by means of electroporation with these plasmids have also been reported [51]. More so, in some of the successful transformation by electroporation using pBBR1MCS-2 backbone, the plasmids were first transferred by conjugation into the bacterium, isolated and then re-transformed by electroporation into the bacterium [16,56].

In this study, it is established that transformation of *C. necator* by electroporation using pBBR1MCS-2 is possible under low concentration of selective pressure, kanamycin. Nevertheless, the transformation efficiencies obtained were very low. Despite the inefficacious transformation of *C. necator* with pBBR1MCS-2 by electroporation, most studies involving the bacterium is carried out with this plasmid backbone and the plasmid delivered by conjugation [12,59,63,88,108,139]. This is likely due to the presence of MCS in the pBBR1MCS-2 in comparison to pBHR1. The broad host range plasmid, pBHR1, has proven instrumental in studies involving *C. necator*, owing to its ability to be electroporated into the bacterium [11,20,53,55,67,73]. This study demonstrates that the ineffective (low efficiency and inconsistent) transformation of *C. necator* with pBBR1MCS-2 is due to the *KanR* in the plasmid. Both pBBR1MCS-2 and pBHR1 have *KanR* for selecting *C. necator* transformants. Nonetheless, their individual *KanR* sequence differs significantly. The length of amino acids (a.a) in pBHR1 KanR is 816 a.a, whilst that in pBBR1MCS-2 is 795 a.a. The codon fractions for both sequences are different (Appendix B). However, pBBR1MCS-2 KanR has codon fraction similar to that of *C. necator*. This indicates that codon difference (bias or usage) might not be responsible for the ineffective transformation of *C. necator* with pBBR1MCS-2. The fact that both *KanR* differ in both protein and DNA sequence implied that they code different class of enzyme, which differ in enzyme kinetics. The *KanR* in pBHR1 codes for aminoglycoside *O*-phosphotransferase class I, subtype 'a' (*APH (3')-Ia*), whilst that in pBBR1MCS-2 codes for (*APH (3')-IIa*). Thus, they belong to the same enzyme subtype, but to a different class. It appears *APH (3')-Ia* is a highly evolved catalyst with broad aminoglycoside substrate specificity [140]. On the other hand, *APH (3')-IIa* is a well-characterised and mapped neo gene, which is derived from the Tn5 transposon [141]. It is widely used in molecular biology as kanamycin and neomycin resistant marker for prokaryotes [142]. Therefore, the low transformation efficiency, inconsistent and unsuccessful electroporation of *C. necator* with

pBBR1MCS-2 is attributed to the *KanR*. It is plausible that within the short period of recovery following electroporation, *APH (3')-IIa* is insufficiently expressed. Thus *C. necator* transformed with plasmid coding for this enzyme could not tolerate high concentration of kanamycin (200 *μ*g/mL), as a result failed to form colonies when selected at this concentration.

Furthermore, this study shows that not all antibiotics are suitable for selecting *C. necator* transformants. The bacterium exhibited high degree of resistance to gentamicin and spectinomycin. *C. necator* H16 is designated gentamicin resistant [12,44]. Due to the degree of resistance to gentamicin, it is more likely that acquisition of one of the aminoglycoside modifying enzymes, aminoglycoside acetyltransferase, by lateral gene transfer is responsible for such resistance [143,144]. Spectinomycin on the other hand, is an aminocyclitol antibiotic, which is mostly rendered ineffective by bacterial adenylyltransferase [145]. Both antibiotics are protein synthesis inhibitors, specifically binding to the 30S subunit of bacterial ribosome to inhibit protein synthesis. Further genetic studies on the bacterium will shed light on the biochemical mechanisms of resistance to these two antibiotics and will potentially offer more antibiotic choices for selecting *C. necator* transformants.

Although kanamycin is often used as a selective pressure for *C. necator* transformants, the antibiotic showed weak antimicrobial activity towards the bacterium. Kanamycin like gentamicin is an aminoglycoside, however, resistance towards it is mostly mediated by 3'-aminoglycoside phosphotransferases, specifically type I and II, which are commonly associated with Gram-negative bacteria. Kanamycin inhibits bacteria growth by binding to the 30S of the ribosome to cause mRNA misreading, resulting in the synthesis of non-functional proteins [146]. *Cupriavidus* is highly susceptible to tetracycline, chloramphenicol and erythromycin [147]. Tetracycline is a broad-spectrum antibiotic, inhibiting not only Gram-negative and -positive bacteria, but also mycobacteria and protozoans [148–150]. This antibiotic binds to the 30S subunit of the ribosome, resulting in a selective and reversible inhibition of proteins synthesis. The high degree of *C. necator* susceptibility to tetracycline, despite its bacteriostatic effect, is ascribed to active concentration of the antibiotic in the cells [148]. Magnesium enhances this active accumulation by facilitating transport across the cell to the target site [149]. This implies that selecting tetracycline transformants will be more effective in magnesium-free growth medium. An earlier study reported increased susceptibility of Gram-negative bacteria, *Pseudomonas* spp., to tetracycline in a culture medium supplemented with magnesium [151]. Erythromycin is also a bacteriostatic protein synthesis inhibitor, with more antimicrobial activity against Gram-negative bacteria. It binds to the 50S subunit of bacterial ribosome and inhibit the elongation stage of protein synthesis [152]. Similarly, chloramphenicol

is a broad-spectrum bacteriostatic protein synthesis inhibitor. However, unlike erythromycin, which blocks post-translocation stage of protein synthesis, chloramphenicol directly binds to substrate residues in the 50S ribosome to prevent peptide bond formation, pre-translocation, by inhibiting the action of peptidyltransferase [153]. Conversely, ampicillin and carbenicillin are inhibitors of cell wall synthesis, selectively inhibiting the function of transpeptidase. This prevents crosslinking of peptidoglycan, leading to cell death by lysis [154]. Although both antibiotics are known to exert bactericidal effect, *C. necator* showed low susceptibility to both antibiotics under shake flask cultivation. Specifically, ampicillin showed variable antimicrobial activity against the bacterium. This effect is ascribed to instability of ampicillin, making the antibiotic not a first choice for selecting *C. necator* transformants.

3.5 Conclusion

Bioparts of widely used broad host range plasmids were characterised for application in *C. necator*. pBHR1 is more efficiently delivered to *C. necator* by electroporation in comparison to pBBR1MCS-2 and pKT230. The ineffective transformation of *C. necator* with pBBR1MCS-2, the widely used plasmid backbone for the bacterium, is due to the *KanR* coded by the plasmid. *Mob* sequence is not essential for transforming *C. necator* by electroporation. Not all antibiotic cassettes are suitable for selecting *C. necator* transformants. Kanamycin, chloramphenicol, tetracycline, erythromycin and carbenicillin cassettes can potentially be used for future applications in *C. necator*. Gentamicin, spectinomycin and ampicillin are not considered suitable as a potential selective pressure for *C. necator*. This study showed that heat-shock transformation of *C. necator* is possible. Components of transformation buffer, heat-shock temperature and time, and the amount of plasmid delivered are important factors to consider when optimising this novel transformation protocol for *C. necator*. Unlike electroporation, which requires less amount of DNA (~50 ng) to achieve high transformation efficiency, heat-shock transformation of *C. necator* would require higher amount of DNA (> 500 ng) to achieve considerable transformation efficiency. Therefore, plasmid backbone is a critical factor to consider when optimising heat-shock transformation protocol. The information obtained from the characterisation of plasmid bioparts will guide the construction of modular minimal plasmids for *C. necator*.

Chapter 4.

Construction of Modular Minimal Plasmids for *Cupriavidus necator* H16

4.1 Introduction

In synthetic biology, plasmids are principal to *in vitro* and *in vivo* genetic manipulation of microorganisms. Most activities in this field of study rely on the use of well-defined and characterised biological parts (bioparts), that can be swiftly exchanged in a modular format, to predict the output of a system. This allows bioengineers to rapidly design, build and test increasingly predictable devices efficiently, with higher repeatability and reproducibility. There are range of modular plasmids for use in bacteria: pZ [155], BioBricks [156], pBAM1 [157] and pSEVA [158]. Although modular, most of these plasmids were constructed with narrow host range replication origin, and the activity range of the bioparts characterised in *E. coli*. This limits the applications of some of these plasmids to other industrially relevant microbial chassis like *C. necator*. The standard European Vector Architecture (pSEVA) format provides a platform for modular bioparts for other microbial chassis to be readily accessible [158]. However, the plethora of high-throughput, efficient and seamless DNA assembly methods have gained significant attention over modularity of plasmid bioparts. These DNA assembly methods BioBricks [159], Gibson [160], Golden gate [161,162], BglBrick [163], SLIC [164] and PaperClip [165] are found useful in rapid construction of customised plasmids for use in bacterial chassis. Golden gate relies on the use of Type IIs restriction enzymes, which recognises non-palindromic sequences and cut at a defined distance outside of the recognition site, to seamlessly and efficiently assemble multiple DNA fragments in a one pot-one step restriction-ligation reaction [162,166–168]. This approach is germane in synthetic biology and metabolic engineering applications, for assembling multi-genes fragments in a pre-determined sequence.

Application of high-throughput DNA assembly to the construction of plasmid intended for propagation in *C. necator* is hindered by several factors, notably plasmid choice and methods of transformation. To transform *C. necator*, a broad host range (BHR) plasmid is required. Among the existing BHR plasmids for *C. necator*, pBHR1 and pBBR1MCS-2 are widely used. Compared with other BHR plasmids, pBHR1 and pBBR1MCS-2 replication sequence (pBBR1) are smaller in size—with the replication protein located together with the replication origin [47–49,135]. Whilst pBHR1 can be delivered by electroporation [11,73,169], plasmids

bearing pBBR1MCS-2 backbone are predominantly delivered by conjugation, with *E coli* S17-1 as the donor bacterium [12,59,88,108]. Despite the drawbacks of conjugation, most of the cloning in *C. necator* are carried out using pBBR1MCS-2, perhaps due to the plasmid having complete and more desirable bioparts. Nevertheless, the backbones of the existing BHR plasmids are not desirable for high-throughput cloning. More importantly, they are not characterised, standardised and modularised for *C. necator*. Additionally, the transformation efficiencies and segregational stabilities obtained with the existing BHR plasmids are considerably low [46,56,59]. These further impede the potential bioengineering applications in *C. necator*.

In chapter three, mobilisation region (*Mob*) was characterised for its function on *C. necator* transformation by electroporation, while the antimicrobial susceptibility of *C. necator* to range of antibiotics was determined. It was established that removal of *Mob* did not affect transformation efficiency following electroporation. With the knowledge that pBHR1—with small replication sequence—is delivered to *C. necator* by electroporation with 100% efficiency, this study is designed to achieve two objectives: first, construct modular minimal plasmids with fully-define sequences, predictable and improved function; and secondly, deliver the resulting plasmids by electroporation. In this chapter, the hypothesis that reduction in plasmid size would improve plasmid transformation efficiency and segregational stability is tested.

4.2 Materials and Methods

4.2.1 Media and molecular reagents

Plasmids harbouring colE1 origins of replication, pBW213ara-hrps (pBR322) and pUC19, together with pBW115lac-hrpR (p15A) were obtained from Addgene. Primers were designed using Primer3 Plus and synthesised by Invitrogen, Life Technologies. Double stranded DNA (gBlock) fragments were synthesised by Integrated DNA Technologies (IDT). *C. necator* chemically defined medium contained fructose (0.5 % w/v), NaH_2PO_4 (0.01% w/v), Na_2HPO_4 (0.01% w/v), K_2SO_4 (0.13% w/v), $MgSO_4$ (0.001% w/v), $CaCl_2$ (0.01% w/v), NH_4Cl (0.001% w/v), trace solution (0.1% v/v) and amino acids solution (1% v/v) Chapter 2. Trace element solution contained (g/L): 15 $FeSO_4.7H_2O$, 2.4 $MnSO_4.H_2O$, 2.4 $ZnSO_4.7H_2O$, and 0.48 $CuSO_4.5H_2O$. Amino acid solution contained (g/L): 12.9 arginine, 10 each of histidine, leucine and methionine. Other media, chemicals and reagents are as described in Chapter 3.2.1. PCR, purification and quantification of PCR products and plasmids were carried out as described previously Chapter 3.2.2. BsaI restriction enzyme was obtained from NEB and T4 DNA ligase

from Promega. All constructs were verified by Sanger DNA sequencing. Primers used for amplification of each biopart are shown in Table 4.1

4.2.2 SEVA design

To build modular plasmids, the design-build-test-learn approach was adopted. The design was based on the SEVA format [158,170]. Unusual (rare-cutter) restriction enzyme sequences were included at the 5' end of each biopart; these set the boundaries for each of the biopart in the final plasmid backbone (Fig. 4.1). Bioparts were amplified from the appropriate plasmid using a Q5® High-Fidelity DNA polymerase (NEB UK). The thermocycling conditions were: 98 °C for 30 s; 30 cycles of 98 °C for 10 s (denaturation), 68 °C for 30 s (annealing), 72 °C for 30 s (elongation), and a final extension at 72 °C for 2 min. Following the amplification, amplicons were purified using a QIAquick PCR purification kit and were further treated with DpnI (NEB UK) to digest the methylated templates (backbone vectors). DpnI digestion was performed at 37 °C for 1 h, followed by 20 min incubation at 80°C. To improve efficiency during the assembly reaction, DpnI digestion reactions were purified using a QIAquick PCR purification kit prior to assembly. The purified digestion products were quantified using a Qubit™ dsDNA high sensitivity (HS) assay kit.

Table 4. 1 Primers for amplification of bioparts.

Biopart	Source	Annealing oligonucleotides
MCS	pBBR1MCS-2	F: 5'-CTCATCGCAGTCGGCCTATT-3' R: 5'-CACTCATTAGGCACCCCAGG-3'
Reporters	Synthesized by IDT	Not applicable
CmR	pBHR1	F: 5'-TTAACGACCCTGCCCTGAAC-3' R: 5'-GGTGTCCCTGTTGATACCGG-3'
KanR	pBHR1	F: 5'-GTCCCGTCAAGTCAGCGTAA-3' R: 5'-GGAAAGCCACGTTGTGTCTC-3'
TcR	pT18mobScab	F: 5'-GTTGGGAAGCCCTGCAAAGT-3' R: 5'-GCCACAGTCGATGAATCCAG-3'
Rep	pBHR1	F: 5'-ATAATTGTTGTCGCGCTGCC-3' R: 5'-CAGACAAGGTATAGGGCGGC-3'
Oriv-Rep	pBBR1MCS-2	F1: 5'-CTTCGCAAAGTCGTGACCGC-3' R1: 5'-CAGACAAGGTATAGGGCGGC-3'
Oriv	pBBR1MCS-2	F1: 5'-CTTCGCAAAGTCGTGACCGC-3' R2: 5'-GCTTATCTCCATGCGGTAGG-3'
Rep	pBBR1MCS-2	F2: 5'-ATGGCCACGCAGTCCAGAGA-3' R1: 5'-CAGACAAGGTATAGGGCGGC-3'
pMB1(*Rop_ori*)	pBR322	F: 5'-TCCAGTAACCGGGCATGTTC-3' R: 5'-TTTCGTTCCACTGAGCGTCA-3'
pMB1 (*ori*)	pUC19	F: 5'-GTGAGCAAAAGGCCAGCAAA-3' R: 5'-GCGTCAGACCCCGTAGAAAA-3'
p15A	pBW115lac_hrpR	F: 5'-CGCTGAGATAGGTGCCTCAC-3' R: 5'-GCGGAAATGGCTTACGAACG-3'

Unusual restriction sequence (rare cutter) sets the boundary between each module; this is added to the 5' end of primer (mostly the forward primers). Upstream of the rare cutter (RC) is a four nucleotide bases (fusion site), with the complementary bases added to the 3' end of the preceding biopart (see Appendix C for detail).

4.2.3 Golden gate assembly (build)

High-throughput assembly of bioparts was accomplished using a Golden gate DNA assembly method [161]. Briefly, ~50 ng of each biopart was combined with 0.75 μL 2 mgL^{-1} molecular biology grade bovine serum albumin (BSA, NEB UK), 1.5 μL 10X ligase buffer (Promega), 1 μL 3 $U\mu L^{-1}$ T4 DNA ligase (Promega) and 2 μL 10 $U\mu L^{-1}$ BsaI (NEB UK) in a one pot-one step restriction-ligation reaction. The total reaction volume was raised to 15 μL with molecular grade water. Subsequently, the reactions were allowed to proceed for 30 cycles of 37 °C for 5 min and 16 °C for 10 min, followed by 50 °C for 5 min and 80 °C for 5 min.

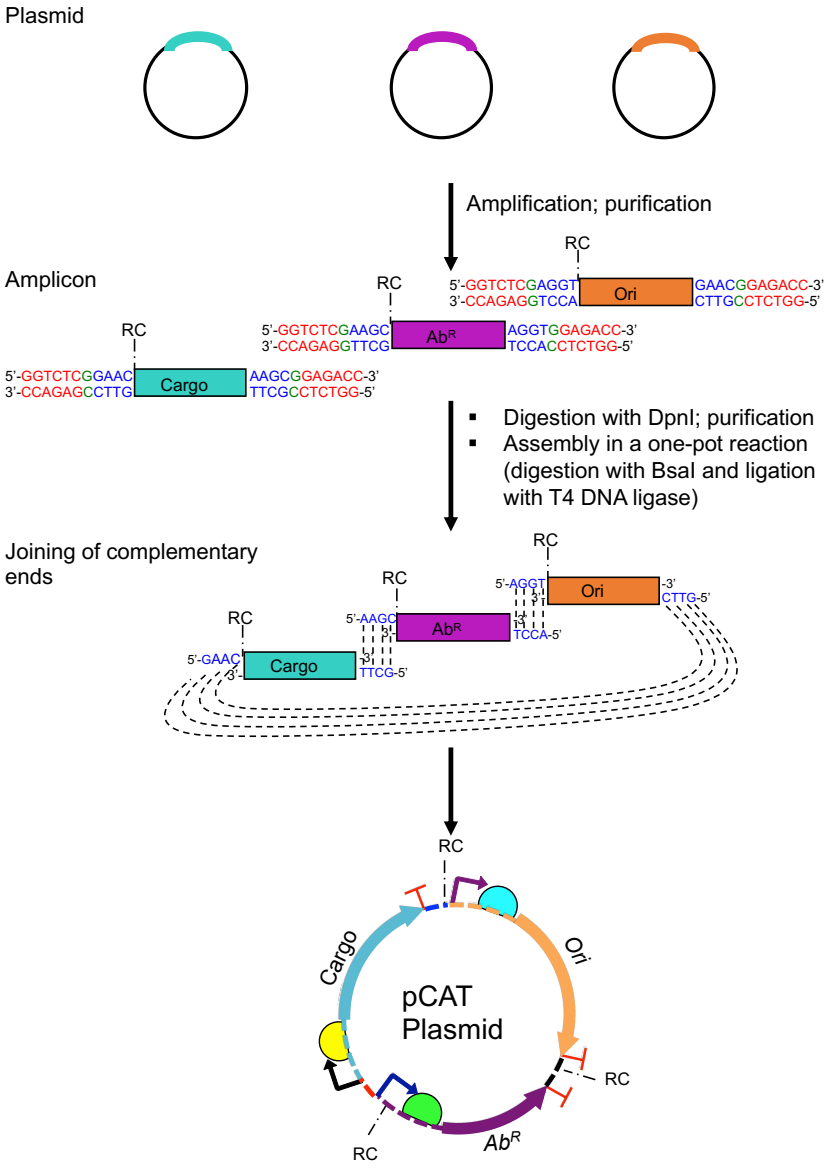


Fig. 4. 1 High-throughput DNA assembly.

Bioparts were first characterised for function in *C. necator* and amplified using a high-fidelity DNA polymerase. PCR fragments (amplicons) were later digested with DpnI to degrade background plasmids and digestion products purified prior to assembly. The assembly was performed in a single reaction tube containing ~ 50 ng of each biopart and other reagents (BsaI, DNA ligase, ligase buffer and bovine serum albumin (BSA)). Bioparts are separated from each other by a rare cutter (RC) restriction sequence (see Appendix C for detail).

4.2.4 Transformation (test)

To test the assembly, 5 µL of the restriction-ligation reaction was combined with 50 µL of high efficiency chemically competent *E. coli* (DH5α) and heat-shocked at 42 °C for 30 s. Transformants were recovered by adding 200 µL of *E. coli* SOC, followed by incubation for 1 h at 37 °C, 220 rpm. After the recovery, 100 µL was spread on LB agar supplemented with suitable antibiotics. Plasmids were recovered from *E. coli* and delivered to *C. necator* by electroporation [Chapter 3.2.2,3.2.3]. To deliver the restriction-ligation reactions directly to *C. necator*,

the reactions products were first purified using a QIAquick PCR purification kit and ~ 50 ng of each reaction electroporated directly to the bacterium.

4.2.5 Reporter gene characterisation

Cupriavidus necator transformed with plasmids bearing reporter genes as the cargo were characterised for the expression of individual reporter protein as described [57,60]. Fresh colonies of transformants were suspended in 50 mL falcon tubes containing 5 mL LB broth supplemented with suitable antibiotics. Tube were incubated for 48 h at 30°C and 200 rpm. Two hundred microliter each was transferred into clear-bottom 96-well microtiter plate. Cell growth, optical density (OD_{600nm}), and protein expression (fluorescence output) were quantified using a Varioskan LUX^{TM} Multimode Microplate reader (Thermo Scientific). The excitation and emission wavelengths (E_x and E_m) for eGFP and mRFP1 were E_x 488 nm, E_m 510 nm; and E_x 584 nm, E_m 607 nm, respectively. The plate reader was set at top read (optics) and 5 nm excitation bandwidth. Relative fluorescence unit (RFU) was determined as the difference in the expression level of strain carrying a reporter to the expression level of a reporter null-strain (control). RFU for each culture was further divided by the corresponding OD_{600nm} to obtain RFU/OD_{600nm}.

4.2.6 Plasmid segregational stability

Plasmid segregational stability was carried out by serial culturing of *C. necator* transformants under selective and non-selective conditions as described [59]. Briefly, a fresh colony of transformant was suspended in 10 μL of molecular grade water in a PCR tube. Five microliter each was transferred into two sets of 50 mL falcon tubes each containing 5 mL fresh LB broth. One of the tubes (control) was made selective by supplementing with suitable antibiotic, whilst the test sample tube was maintained under non-selective condition, without the addition of antibiotic (Fig. 4.2). Both tubes were incubated under identical conditions: 30°C, 200 rpm for 24 h. Subsequently, 5 μL from each 24 h culture tube was transferred into corresponding sets of new tubes, containing fresh broth with or without antibiotic (Fig. 4.2). The freshly inoculated tubes were incubated as described previously. This procedure was repeated every 24 h for up to 144 h for both the control and the test sample tube. Additionally, at every 24 h interval, culture from the test sample tube was diluted 10^{-6}-fold. One hundred microliter of the diluted culture was spread on two sets of LB agar plates: one supplemented with antibiotic and the other without antibiotic (Fig. 4.2). All plates were incubated under appropriate condition. For colonies originating from the test sample tube, percentage segregational stability was determined as the ratio of colony-forming unit (CFU) on agar plate with antibiotic to CFU on agar plate without antibiotics. For the control, grown

under selective conditions, cultures were diluted as with the test sample, spread only on selective LB agar and incubated under the same condition. However, this was performed at two intervals, 48 h and 144 h.

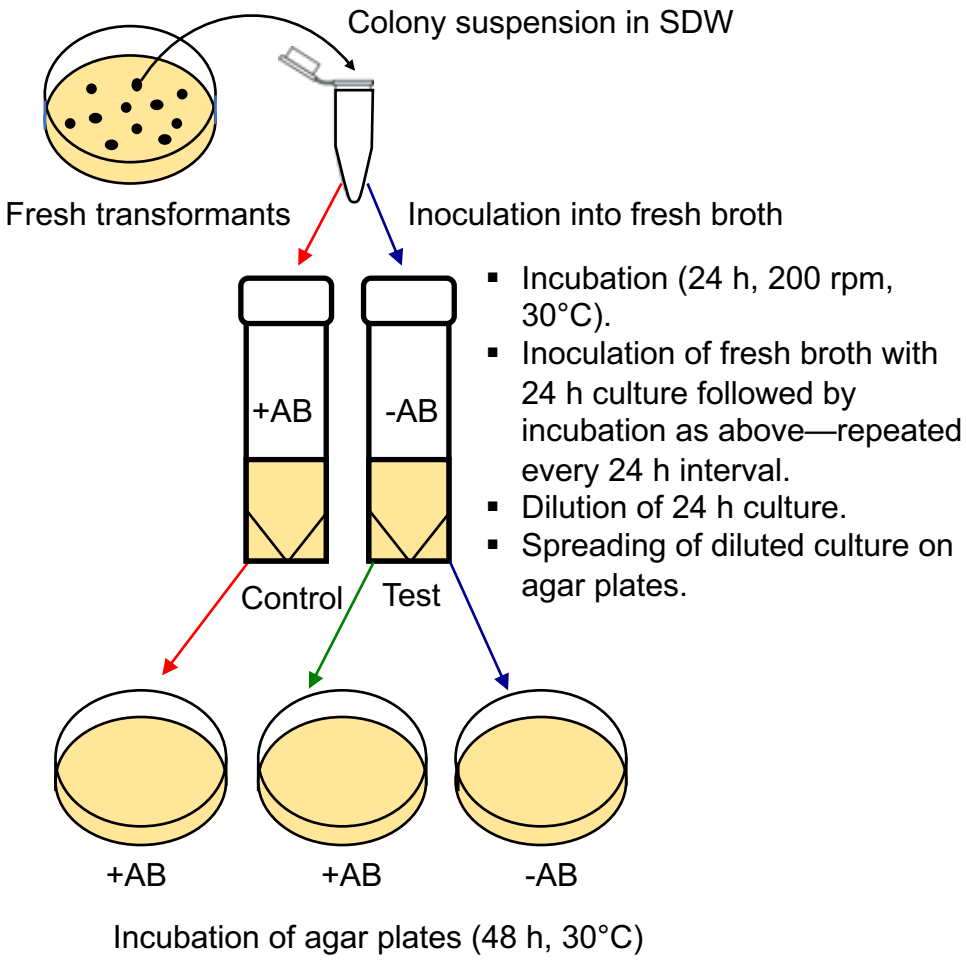


Fig. 4. 2 Determination of plasmid segregational stability.

Fresh *C. necator* transformants were suspended in 10 µL sterile distilled water (SDW), and 5 µL each was transferred into 50 mL falcon tube containing fresh 5 mL LB broth. One of the tubes (+AB) was made selective by supplementing with appropriate antibiotic; the other (-AB) was maintained under non-selective condition. Both tubes were incubated under identical conditions for 24 h. At every 24 h interval, 5 µL culture from each tube was transferred into similar fresh broth and incubated under the same conditions. Also, at this interval, culture was diluted 10^{-6} and spread on appropriate selective and non-selective LB-agar plates. This procedure was carried out for *n* = 3 biological replicates; AB = antibiotic.

4.3 Results

4.3.1 Plasmid design and choice of bioparts for constructing modular minimal plasmids

To construct modular minimal plasmids (pCAT) that can be delivered to *C. necator* by electroporation, plasmids were designed to have bioparts that are essential for replication in *C. necator*. The plasmid design is based on the SEVA format, with each biopart separated by an unusual restriction site (rare cutter). The SEVA plasmid layout has six functional bioparts (modules), three of which are fixed connectors and the other three are variable bioparts (Fig. 4.3). The fixed modules are two terminators flanking the cargo (MCS or reporter protein) and an origin of conjugative transfer (*oriT*). The variable bioparts are origin of replication, cargo and antibiotic resistant cassette (Ab^R). Unlike the SEVA format, the plasmids in this study have fewer bioparts (*Rep*, Ab^R and cargo), which are variable. The layout of the pCAT plasmids are as described (Fig. 4.4). In chapter three, it was demonstrated that *Mob* or *oriT* is dispensable for transformation of *C. necator* by electroporation. Thus, pCAT plasmid design excludes this fixed biopart in the SEVA format. The design also excludes the first connector (Terminator, T1) located between the replication machinery and cargo in the SEVA description. The second connector (Terminator, T0) is not obvious in the pCAT design and is located together with the reporter biopart as terminator(s). This small refactoring of the SEVA design improves the flexibility of pCAT plasmids. It further provides a platform for the plasmid biopart to be sourced individually and independently of the plasmid backbone and assembled in a high-throughput method with other bioparts characterised for use in other microbial chassis.

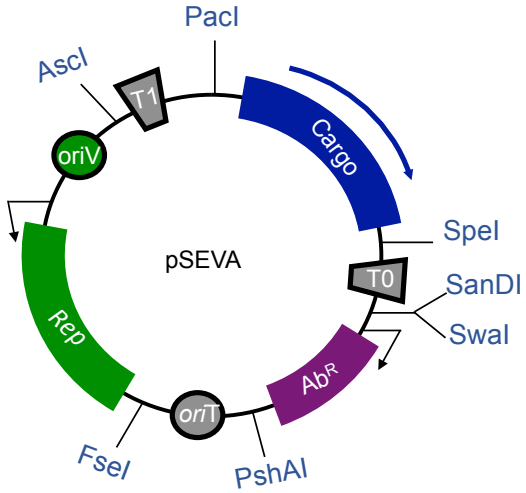


Fig. 4. 3 The Standard European Vector Architecture (SEVA) plasmid layout.

Six functional modules of SEVA plasmid: three variable (cargo, antibiotic resistant marker (Ab^R) and replication sequence (*oriV-Rep*)) module; and three fixed (two terminators (T1 and T0) and origin of transfer *oriT*) modules. Each module is separated by an unusual (rare) restriction enzyme sequence: PacI, SpeI, SanDI, SwaI, PshAI, FseI, and AscI [170].

The choice of pCAT plasmid bioparts is driven by the results of bioparts characterisation in *C. necator*. The plasmid replication sequence (*Rep*) was amplified from pBHR1 since the plasmid has so far been serving as a control. Multiple cloning site (MCS) was amplified from pBBR1MCS-2, whilst reporter proteins were from already characterised reporter cassettes for *C. necator* [57] and another set of cassettes extensively characterised in *E. coli* [60].

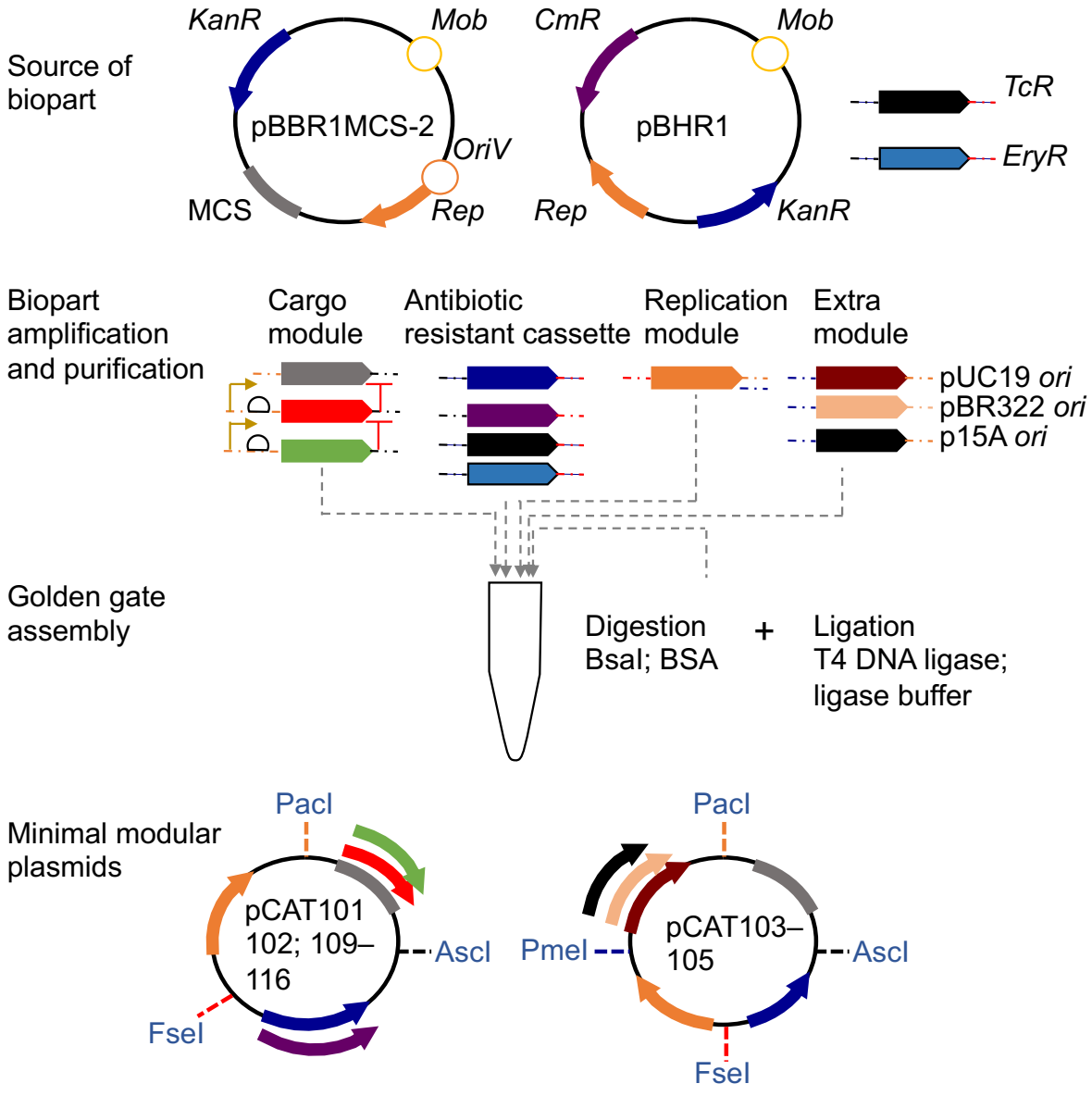


Fig. 4. 4 Construction of modular minimal plasmids.

Replication sequence (*Rep*), kanamycin and chloramphenicol resistant cassettes (*KanR*) and (*CmR*), respectively were amplified from pBHR1. Multiple cloning sites (MCS) was amplified from pBBR1MCS-2. Reporter gene cassettes (mRFP1 and eGFP) were chemically synthesised. Promoters, ribosomal binding sites and terminator sequences are provided in Appendix C. The source of other bioparts are as shown (Table 4.1). Bioparts were assembled in a one pot-one step restriction-ligation reaction. Unusual restriction sites (PacI, AscI, FseI and PmeI) set the boundaries between each biopart. Narrow host range replication sequences (*ori*) formed the fourth module in the four bioparts plasmid layout.

Informed by the knowledge of *C. necator* antimicrobial susceptibility, tetracycline, chloramphenicol, erythromycin and kanamycin resistant cassettes were chosen as potential markers that could be used for selecting *C. necator* transformants. Carbenicillin and ampicillin were excluded due to their bactericidal effect. Additionally, ampicillin was excluded owing to its inconsistent antimicrobial activity against *C. necator*. Antibiotic cassettes: *KanR* (*APH (3')-Ia*) and *CmR* were amplified from pBHR1, *EryR* from pTRKH3 and *TcR* from pT18mobsacB. The source of other bioparts are as shown (Table 4.1).

Table 4. 2 pCAT plasmid and biopart.

Plasmids ID	Description	T.E.	Source
pUC19	**pMB1 *ori*;** MCS; Amp	N.D.	[171]
pBR322	**pMB1 *ori_rop*;** *TcR*; Amp	N.D.	[172]
pBW115lac-hrpR	**p15A**; lacI; *KanR*; *SmR*	N.D.	[173]
pT18mobsacB	*Ori*; *oriT*; MCS; ***TcR***; *SacB*	N.D.	[174]
pTRKH3	pAMβ1 *ori*; p15A *ori*; *TcR*, ***EryR***	N.D.	[175]
pKT230	*nic oriV*; *Mob*; *SmR*; *SuR*	N.D.	[135]
[a]pBHR1	**pBBR1 *Rep;*** *Mob*; ***KanR***; ***CmR***	5.77 x 10^3 ± 31%	[49]
[b]pBHR1	pBBR1 *Rep*; *Mob*; *KanR*; *CmR*	6.50 x 10^2 ± 23%	[49]
pBBR1MCS-2	pBBR1 *oriV-Rep*; *Mob*; *KanR*; **MCS**	1.35 x 10^3 ± 14%	[48]
pBBR1MSC-2a	pBBR1 *oriV-rep*; *Mob*; *KanR*; MCS	1.0 x 10^4 ± 6%	This study
pCAT101	pBBR1 *Rep*; MCS; *KanR*	3.24 x 10^3 ± 34%	This study
pCAT102	pBBR1 *Rep*; MCS; *CmR*	4.30 x 10^3 ± 17%	This study
pCAT103	pBBR1 *Rep*; MCS; *KanR*; pUC19 *ori*	3.95 x 10^4 ± 1%	This study
pCAT104	pBBR1 *Rep*; MCS; *KanR*; pBR322 *ori*	3.70 x 10^4 ± 6%	This study
pCAT105	pBBR1 *Rep*; MCS; *KanR*; p15A *ori*	1.32 x 10^5 ± 6%	This study
pCAT106	pUC19 *ori*; MCS; *KanR*	N.D.	This study
pCAT107	pBR322 *ori;* MCS; *KanR*	N.D.	This study
pCAT108	p15A *ori*; MCS; *KanR*	N.D.	This study
pCAT109	pBBR1 *Rep*; J23100_eGFP; *KanR*	N.D.	This study
pCAT110	pBBR1 *Rep*; J23100_mRFP; *KanR*	N.D.	This study
pCAT111	pBBR1 *Rep*; P_{j5}_eGFP; *KanR*	N.D.	This study
pCAT112	pBBR1 *Rep*; P_{j5}_mRFP; *CmR*	N.D.	This study
pCAT113	pBBR1 *Rep*; $P_{j5[C2]}$_mRFP; *CmR*	N.D.	This study
pCAT114	pBBR1 *Rep*; P_{g25}_mRFP; *CmR*	N.D.	This study
pCAT115	pBBR1 *Rep*; P_{tac} _eGFP; *KanR*	N.D.	This study
pCAT116	pBBR1 *Rep*; P_{tac} _mRFP1; *KanR*	N.D.	This study

Bold bioparts are the parts sourced from each corresponding plasmid. All transformation efficiencies (T.E.) were obtained as transformants/μg DNA [Chapter 3.2.3]. pBHR1 is the existing (control) broad host range plasmid delivered to *C. necator* and selected under kanamycin pressure ([a]pBHR1) or chloramphenicol pressure ([b]pBHR1). pBBR1MCS-2a is a hybrid plasmid derived by substituting *KanR* from pBHR1 with that in pBBR1MCS-2. The transformation efficiencies of plasmids that were unable to transform *C. necator* are designated not determined (N.D.). Percentage error are S.E.M., *n* = 3 biological replicates.

4.3.2 Modular minimal plasmids with three bioparts

Consequently, plasmids were assembled in a high-throughput method. All restriction-ligation reactions were first delivered to *E. coli*, purified and delivered to the destination host, *C. necator*. Variants of plasmids bearing *KanR* or *CmR* transformed *C. necator* and were designated pCAT101 and 102, respectively (Table 4.2). However, variants carrying *EryR* were unable to transform *C. necator* after repeated attempts (*n* = 3 Golden gate assembly). Similarly, variant with *TcR* was unable to transform *C. necator*—even after selecting on LB-agar plates supplemented with lower concentration of tetracycline (2.5 μg/mL). Moreover, this variant ineffectively transformed *E. coli*; the best result yielded three *E. coli* transformants after repeated attempts (*n* = 6 Golden gate assembly). Unlike the *TcR* plasmid variant, *EryR* variants resulted in some *C. necator* colonies, which were later confirmed to be non-transformants by a colony PCR. This growth of non-transformed *C. necator* colonies at suitable erythromycin concentration renders *EryR* unsuitable for selecting *C. necator* transformants. Although no *C. necator* transformant was recovered for variant with *TcR*, tetracycline was still considered a potential selective pressure for the bacterium.

Cognisant that there are two *KanR* that are inadvertently used for selecting *C. necator*, the hypothesis that one of the cassettes—*KanR* from pBBR1MCS-2 coding for *APH (3')-IIa*—is unable to code sufficient kanamycin resistance following electroporation was further substantiated. Therefore, plasmid carrying this *KanR*, *APH (3')-IIa* was assembled. As expected, the resulting plasmid transformed *E. coli* but was unable to transform *C. necator* after repeated attempts (*n* = 3 Golden gate assembly). Hence, it is evident that the ineffective transformation of *C. necator* with pBBR1MCS-2 by electroporation is due to the antibiotic resistant cassette.

4.3.3 Modular minimal plasmids with four bioparts

One of the major challenges encountered so far is low plasmid yield (< 10 ng/μL) following plasmid recovery from *E. coli*. This, perhaps, is not surprising as pBBR1 is a low or medium copy replicon. The low plasmid yield was obtained for plasmids with pBBR1 replication sequence: pBHR1, pBBR1MCS-2 and the modular minimal plasmids, pCAT. The low plasmid yield exacerbates when kanamycin is the selective pressure in comparison to chloramphenicol. To circumvent this, plasmid backbone was extended to four bioparts by including *E. coli ori*: pMB1 (pUC19 and pBR322) and p15A (Fig. 4.4). Plasmids with *E. coli* replication origins transformed *C. necator* with high transformation efficiencies; the plasmids were designated pCAT103–105 (Table 4.2). As expected, only that bearing pBBR1 *Rep* and pUC19 *ori* (pCAT103) gave a very high plasmid yield, whilst variants with pBBR1 *Rep* and

pBR322 or p15A (pCAT104 and 105, respectively) gave low yield (Fig. 4.5A). It is noteworthy that pCAT103 gave inconsistent yields. When the yield was very high (> 100 ng/μL), the resulting plasmid was unable to transform *C. necator*. However, when the yield was moderate (< 10 ng/μL), the resulting plasmid successfully transformed *C. necator* (Fig. 4.5B).

Due to the variability in pCAT103 yield and significant increase in the transformation efficiencies of pCAT103–105, it was unclear which replication sequence, pMB1, p15A or pBBR1 *Rep*, directed replication in *C. necator*. To gain understanding of this, variants with these narrow host range *ori* (pMB1 and p15A) and in turn lacking pBBR1 *Rep* were assembled. Similarly, variant with pBR322 and p15A (pCAT107 and 108) gave low plasmid yield (< 10 ng/μL). Further, unlike pCAT103, pCAT106 with only pUC19 *ori* gave consistent high plasmid yield (> 150 ng/μL). Nonetheless, none of these plasmid variants (pCAT106–108) transformed *C. necator* (Table 4.2). Thus, it was established that pBBR1 *Rep* directed replication of pCAT103–105 in *C. necator*. It appears that inclusion of a high copy *ori* to plasmid backbone that is intended for propagation in *C. necator* might affect the establishment of such plasmid in the bacterium, especially when the plasmid yield is significantly high ($p < 0.05$). Consequently, plasmids (pCAT109–116)—carrying reporter proteins as the cargo—were assembled using only pBBR1 *Rep* sequence to direct replication in both *E. coli* and *C. necator*.

4.3.4 Characterisation of reporter gene

To validate the suitability of plasmids in expressing foreign gene(s) in *C. necator*, pCAT109–116 were characterised for the expression of their individual reporter proteins. These plasmids successfully transformed *E. coli* and expressed their individual reporter in *E. coli*. However, plasmids recovered from *E. coli* colonies expressing reporter proteins on agar plates were unusually high. The resulting plasmids with significant increase in yield were unable to transform *C. necator* after repeated attempts ($n > 3$ biological replicates) (Table 4.2). Conversely, plasmids recovered from *E. coli* colonies expressing no visible reporter protein on agar plates were low. Such plasmids with low yields transformed *C. necator* but failed to express the corresponding cargo reporter protein in the bacterium (Fig. 4.5C). The observation that plasmids with high yields are unable to transform *C. necator* was made previously for pCAT103 (Fig. 4.5A and B). Therefore, the unsuccessful transformation of *C. necator* with pCAT109–116 was ascribed to high plasmid yield rather than reporter cassettes (Fig. 4.5C).

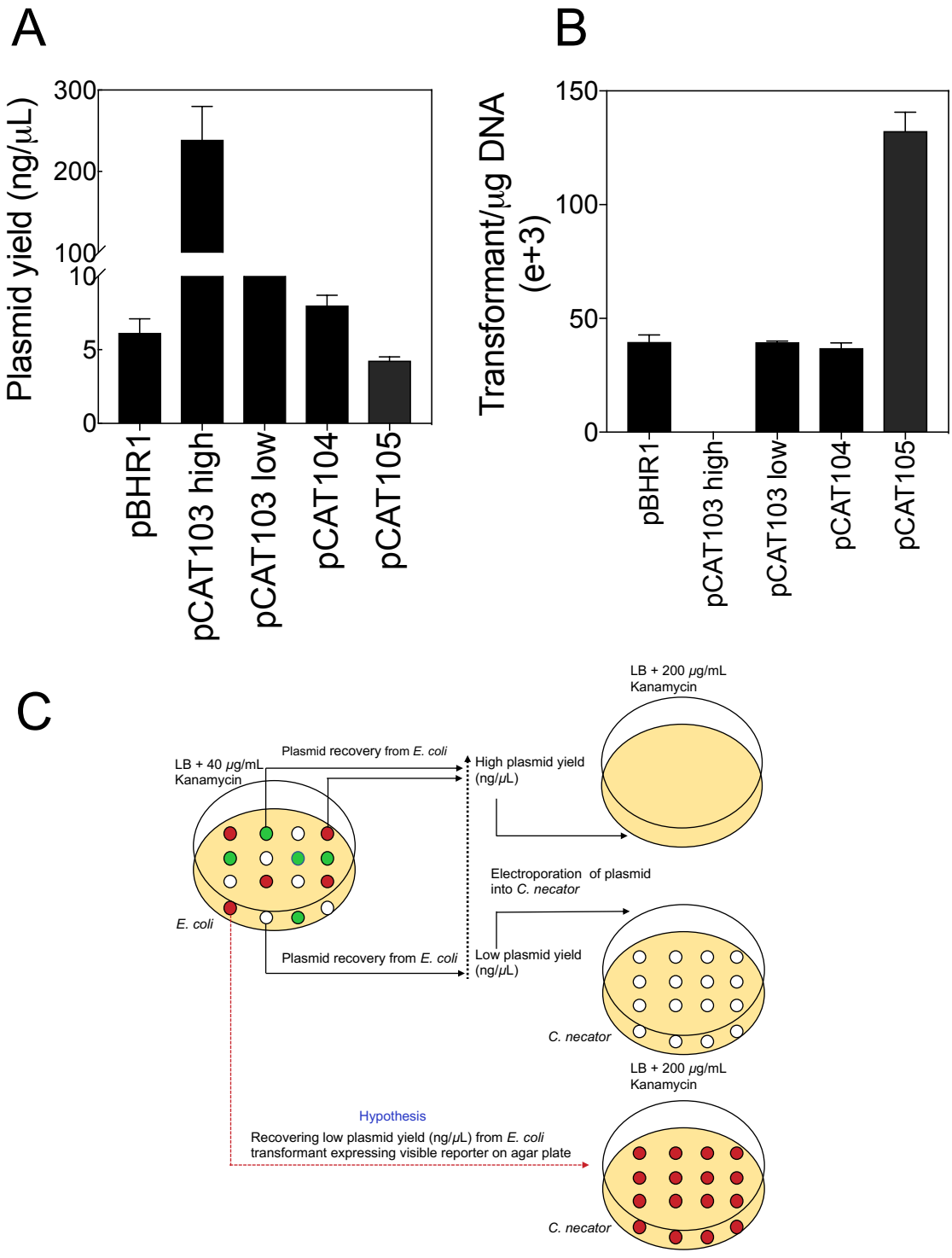


Fig. 4. 5 Effect of plasmid yield on *C. necator* transformation efficiency.

A. Effect of colE1 *ori* on plasmid yields; significant increase in plasmid yield was observed with addition of pUC19 *ori* (pCAT103). For each plasmid, 1.5 mL overnight culture was isolated and eluted in 50 µL molecular grade water. **B.** *C. necator* transformation efficiencies following transformation with plasmids bearing additional *ori*. Plasmid with high yield were not well propagated in *C. necator*. Error bars are S.E.M, *n* = 3 biological replicates. **C.** Summary of the effect of plasmid yield on *C. nectaor* transformation efficiency. Plasmid carry pBHR1 Rep as the replication sequence, *KanR* as the antibiotic cassette and reporter proteins as the cargo. Transformants expressing no visible reporter (white circles), those expressing visible mRFP1 (red circles) and those expressing eGFP (green circles) were tested. High yield (> 100 ng/µL), low yield (< 10 ng/µL).

At this stage, it appears that there is a link between recovering high plasmid yield from *E. coli* and low or no transformation of *C. necator*. Several attempts were made to obtain low plasmid yield from *E. coli* colonies expressing reporter protein on agar plate. Further, different strains of *E. coli* (dam-/dcm- and TOP10) were transformed with fresh restriction-ligation reactions in order to recover plasmid with low yield from colonies expressing reporter protein on agar plate. None of the approaches yielded the intended results: low plasmid yield from *E. coli* colonies expressing reporter proteins on agar plate, which in turn will result in transformation of *C. necator* and subsequent expression of the reporter(s) in *C. necator*. Moreover, restriction-ligation reaction was delivered directly to *C. necator* and the resulting transformants did not express the cloned reporter. The transformation efficiency obtained for *C. necator* via this approach was significantly low ($4.62 \times 10^2 \pm 10\%$ transformants/μg DNA). It is likely that a biopart in the plasmids is unable to regulate and maintain stable expression of foreign genes (reporter proteins) in *C. necator*.

Therefore, it was hypothesised that if *E. coli* colonies expressing visible reporter protein gave low plasmid yield, the resulting plasmid will successfully transform *C. necator* and the cloned gene expressed in the bacterium (Fig. 4.5C).

4.3.5 Recognising the difference between pBHR1 and pBBR1MCS-2 replication sequence

With the understanding that recovering a plasmid with low yield from *E. coli* colonies expressing reporter protein on agar plate will result in subsequent expression of the reporter in *C. necator*, each biopart in the pCAT plasmid was re-examined. Given that the plasmids under investigation have three bioparts, one of which has been characterised (antibiotic resistant cassette), the other (cargo) currently under investigation, attention was focussed on the third biopart, the replication sequence. Previously, the difference between pBHR1 and pBBR1MCS-2 replication sequences was highlighted (Fig. 4.4). It appears pBHR1 has only a replication protein (Rep) directing autonomous replication across bacteria. This pBBR1 *Rep* also directs replication of pCAT plasmids. Unlike pBHR1, pBBR1MCS-2 has both *oriV* and *Rep* directing its replication. The *Rep* of both plasmids is identical; however, in pBHR1 the sequence is preceded by a partial *oriV* sequence (Fig. 4.6). Hence, another set of plasmids carrying the complete broad host range replication machinery (pBBR1 *oriV* and *Rep*) from the pBBR1MCS-2 were assembled (Fig. 4.6).

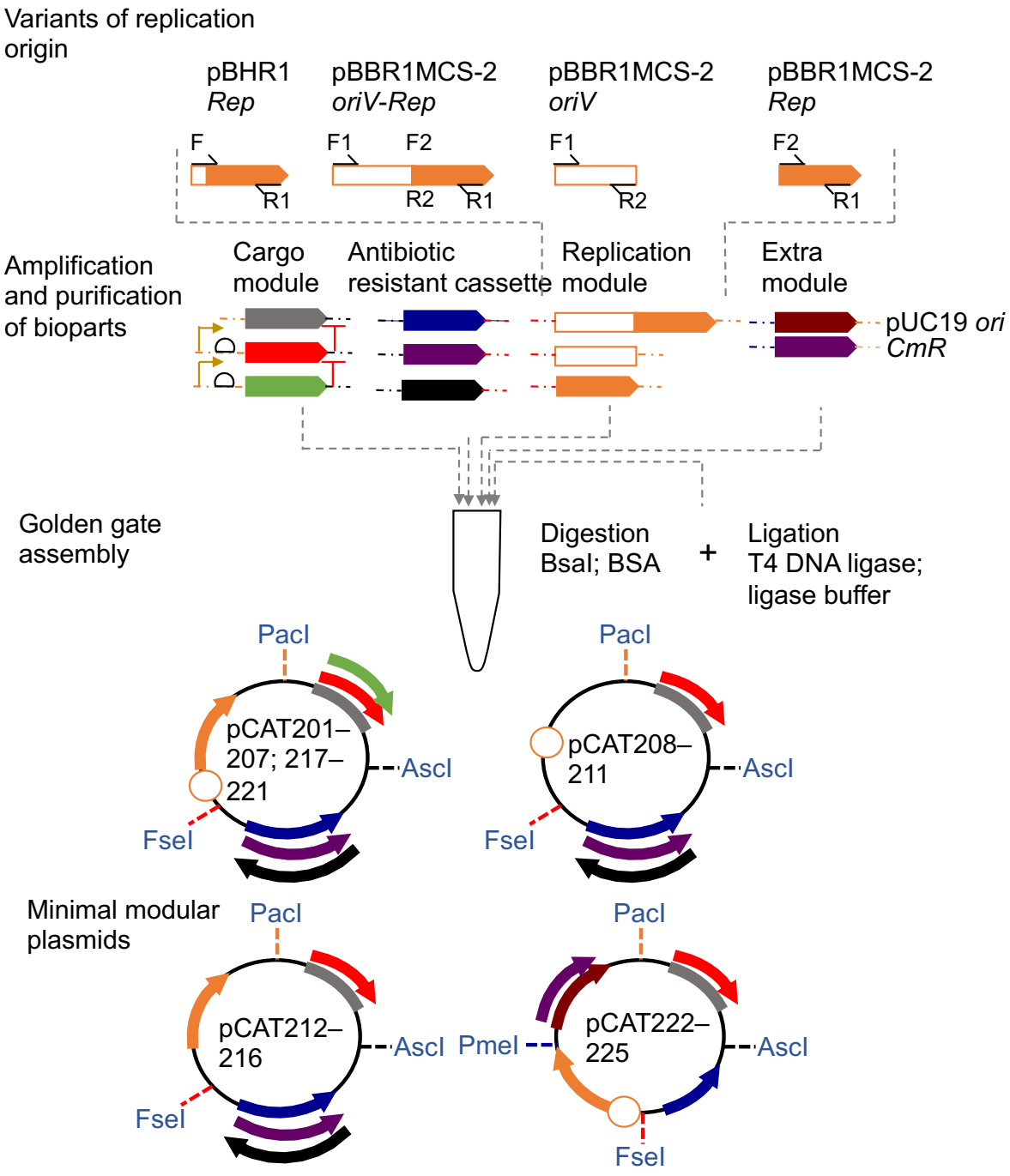


Fig. 4. 6 Plasmid assembly with complete replication sequence.

Complete replication sequence (*oriV-Rep*), partial replication sequence *oriV* or *Rep* and multiple cloning sites (MCS) were amplified from pBBR1MCS-2. Kanamycin and chloramphenicol resistant cassettes, *KanR* and *CmR*, respectively were amplified from pBHR1. Reporter gene cassettes (mRFP1 and eGFP) were chemically synthesised. Promoters, ribosomal binding sites and terminator sequences are given in Appendix C. The sources of other bioparts are as shown (Table 4.1). Bioparts were assembled in a one pot-one step restriction-ligation reaction. Unusual restriction sites (PacI, AscI, FseI and PmeI) set the boundaries between each bioparts. pUC19 *ori*, *CmR* or reporter cassette formed the fourth module in the four bioparts plasmid layout. Cargoes and antibiotic resistant cassettes modules are as described for source of biopart in Fig. 4.4.

Remarkably, plasmids recovered from *E. coli* colonies expressing reporter protein on agar plate, from this new set of plasmids, gave low plasmid yields. The resulting plasmids transformed *C. necator* with very high efficiencies, which were 3000- and ~ 5000-folds higher than that obtained for pBHR1 and pCAT101, respectively (Table 4.3). Most importantly, the resulting *C. necator* transformants expressed visible fluorescence (mRFP1) on agar plates (Fig. 4.7). Amongst this newest set of plasmids, it was observed that pCAT203, with a synthetic promoter ($P_{j5[C2]}$), had more than three-fold higher relative fluorescent unit (RFU) in comparison to pCAT202, with reporter protein expression driven by a non-synthetic promoter (P_{j5}) (Fig. 4.8). Additionally, there was no significant difference (p = 0.38) in the expression level of mRFP1 both under selective and non-selective condition (Fig. 4.9). Subsequently, $P_{j5[C2]}$ was used to drive the expression of reporter protein(s). Plasmid variants bearing *CmR*, pCAT205, also yielded *C. necator* transformants expressing high level of mRFP1. Unlike earlier assembly with only pBBR1 *Rep*, variants bearing *TcR* were also successfully assembled (pCAT206 and 207). These variants gave high transformation efficiencies in *E. coli* and subsequently transformed *C. necator* (Table 4.3). Therefore, the hypothesis that plasmids recovered with high yields from *E. coli* impairs *C. necator* transformation holds true.

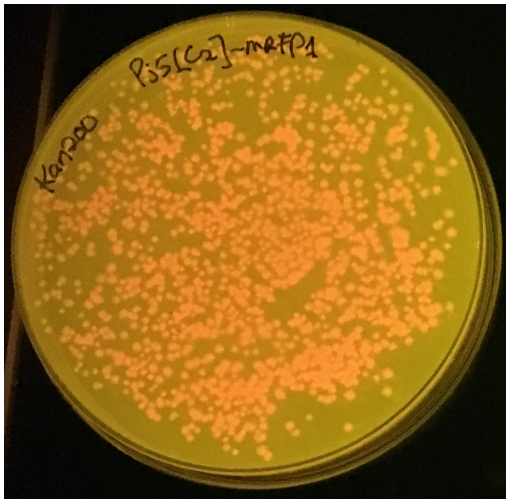


Fig. 4. 7 *Cupriavidus necator* colonies expressing mRFP1 on LB-agar plate after 48 h incubation.

Table 4. 3 Modular minimal plasmids with complete replication sequence.

Plasmids ID	Addgene ID	Description	Size (bp)	T.E.	Sources
pCAT201	134878	pBBR1 *oriV-Rep*; MCS; *KanR*	3198	$1.70 \times 10^7 \pm 14\%$	This study
pCAT202	134879	pBBR1 *oriV-Rep*; P_{j5}_mRFP1; *KanR*	3557	$1.79 \times 10^7 \pm 19\%$	This study
pCAT203	134880	pBBR1 *oriV-Rep*; $P_{j5[C2]}$_mRFP1; *KanR*	3583	$1.90 \times 10^7 \pm 46\%$	This study
pCAT204	134881	pBBR1 *oriV-Rep*; MCS; *CmR*	3161	$5.40 \times 10^5 \pm 35\%$	This study
pCAT205	134882	pBBR1 *oriV-Rep*; $P_{j5[C2]}$_mRFP1; *CmR*	3546	$3.50 \times 10^5 \pm 59\%$	This study
pCAT206	134883	pBBR1 *oriV-Rep*; MCS; *TcR*	3766	$1.18 \times 10^3 \pm 53\%$	This study
pCAT207	134884	pBBR1 *oriV-Rep*; $P_{j5[C2]}$_mRFP1; *TcR*	4151	$1.07 \times 10^3 \pm 62\%$	This study
pCAT208	N.D.	pBBR1 *oriV*; MCS; *KanR*	2454	$5.65 \times 10^3 \pm 12\%$	This study
pCAT209	N.D.	pBBR1 *oriV*; $P_{j5[C2]}$_mRFP1; *KanR*	2839	$6.20 \times 10^3 \pm 15\%$	This study
pCAT210	N.D.	pBBR1 *oriV*; MCS; *CmR*	2416	$4.43 \times 10^5 \pm 15\%$	This study
pCAT211	N.D.	pBBR1 *oriV*; $P_{j5[C2]}$_mRFP1; *CmR*	2802	$7.27 \times 10^5 \pm 30\%$	This study
pCAT212	N.D.	pBBR1 *Rep*; MCS; *KanR*	2306	$2.97 \times 10^3 \pm 19\%$	This study
pCAT213	N.D.	pBBR1 *Rep*; $P_{j5[C2]}$_mRFP1; *KanR*	2691	$5.98 \times 10^3 \pm 9\%$	This study
pCAT214	N.D.	pBBR1 *Rep*; $P_{j5[C2]}$_mRFP1; *CmR*	2654	$4.81 \times 10^5 \pm 13\%$	This study
pCAT215	N.D.	pBBR1 *Rep*; MCS; *TcR*	2874	N.D.	This study
pCAT216	N.D.	pBBR1 *Rep*; $P_{j5[C2]}$_mRFP1; *TcR*	3259	N.D.	This study
pCAT217	134885	pBBR1 *oriV-Rep*; $P_{tac[C2]}$_mRFP1; *KanR*	3549	$6.13 \times 10^5 \pm 42\%$	This study
pCAT218	134886	pBBR1 *oriV-Rep*; $P_{tac[C2]}$_eGFP; *KanR*	3564	$3.66 \times 10^7 \pm 17\%$	This study
pCAT219	134887	pBBR1 *oriV-Rep*; P_{j23100}_$_{PETRBS}$_mRFP1; *KanR*	3579	$2.61 \times 10^7 \pm 13\%$	This study
pCAT220	134888	pBBR1 *oriV-Rep*; P_{j23100}_$_{PETRBS}$_eGFP; *KanR*	3598	$1.28 \times 10^6 \pm 47\%$	This study
pCAT221	134889	pBBR1 *oriV-Rep*; $P_{j5[C2]}$_eGFP; *KanR*	3598	$1.45 \times 10^5 \pm 33\%$	This study
pCAT222	134890	pBBR1 *oriV-Rep*; MCS; KanR; *CmR*	4158	$3.39 \times 10^5 \pm 40\%$	This study
pCAT223	N.D.	pBBR1 *oriV-Rep*; $P_{tac[C2]}$_mRFP1; KanR; *CmR*	4509	$1.73 \times 10^5 \pm 38\%$	This study
pCAT224	134892	pBBR1 *oriV-Rep*; pUC19; MCS; *KanR*	3891	$1.48 \times 10^6 \pm 6\%$	This study
pCAT225	N.D.	pBBR1 *oriV-Rep*; pUC19; $P_{tac[C2]}$_ mRFP1; *KanR*	4242	$6.11 \times 10^5 \pm 36\%$	This study

Purified amplicons (bioparts) were assembled using Golden gate assembly method and plasmids extensively characterised in *C. necator* for transformation efficiency and expression of reporter proteins. All transformation efficiencies (T.E.) were obtained as transformants/μg DNA. The transformation efficiencies of plasmids that were unable to transform *C. necator,* or plasmid not deposited with Addgene are designated not determined (N.D). Percentage S.E.M., *n* = 3 biological replicates.

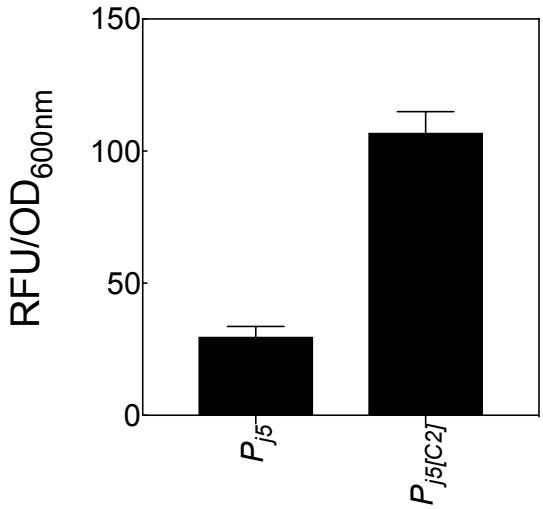


Fig. 4. 8 Comparison of fluorescence (mRFP1) output under synthetic and non-synthetic promoter.

Results shown are for measurement obtained at 48 h incubation. Error bars are S.E.M., *n* = 3 biological replicates.

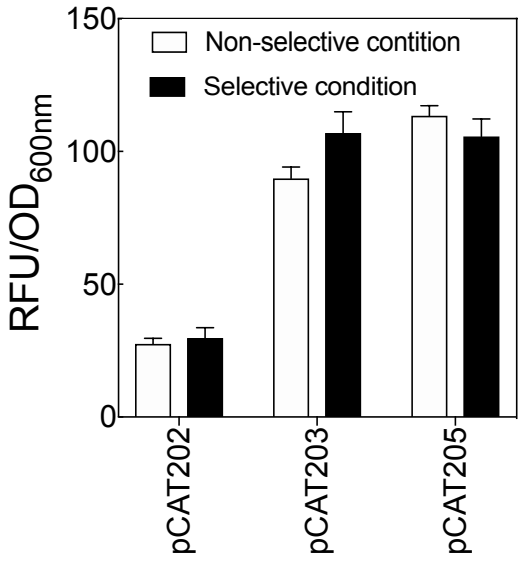


Fig. 4. 9 Expression of mRFP1 under selective and non-selective conditions.

Results shown are for measurement obtained at 48 h incubation. Error bars are S.E.M., *n* = 3 biological replicates.

4.3.6 pBBR1 OriV and pBBR1 Rep independently direct replication in C. necator

It has been demonstrated that the use of complete replication machinery (*oriV-Rep*) results in significant improvement in transformation efficiency. The ability of each replication component to independently direct replication, especially directing the replication and expression of a reporter protein in *C. necator*, was further investigated. Aware that pBHR1

and pBBR1MCS-2 have the same replication protein (pBBR1 Rep) from the same source, *Bordetella bronchiseptica*, the original *Rep* sequence from pBHR1 used in assembling pCAT101–116 plasmid sets were re-examined. In pBHR1, the *Rep* sequence spans 3049–3711 bp, whereas 2964–3792 bp was amplified and used to direct replication in pCAT101–116. It was realised that the upstream sequence of the pBHR1 *Rep* (2642–3048 bp) is the downstream sequence of pBBR1MCS-2 *oriV* (1386–1792 bp) (Fig. 4.6). This indicates that the additional 85 bp (region 2964–3049) upstream of the *Rep* sequence in pCAT101–116 plasmid is a partial sequence of the *oriV* of pBBR1MCS-2. With this understanding, a pair of internal primers were designed to independently amplify the *oriV* and *Rep* sequence—with the primer set used in amplifying the *oriV-Rep* in the pCAT201–207 plasmids defining the external regions (Fig 4.6). The *oriV-Rep* in pCAT201–207 (901–2536 bp in pBBR1MCS-2) were initially amplified with F1: 5'-CTTCGCAAAGTCGTGACCGC-3' and R1: 5'-CAGACAAGGTATAGGGCGGC-3', the same reverse primer used in amplifying the reverse sequence of pBHR1 *Rep* in the pCAT101–116 plasmids (Table 4.1). Thus, the reverse primer defined the same 3' region of the pBBR1 *Rep*. The internal primer set were F2: 5'-ATGGCCACGCAGTCCAGAGA (region 1793–1812) and R2: 5'-GCTTATCTCCATGCGGTAGG-3' (region 1773–1792). With these two pair of internal primers, *oriV* was amplified with F1 and R2, whilst the coding sequence of the *Rep*, with no partial upstream *oriV* sequence, was amplified with F2 and R1 (Fig 4.6).

Both *oriV* and the *Rep* independently directed replication of plasmids in *C. necator*. The plasmids from this assembly were designated pCAT208–214 (Table 4.3). Intriguingly, both were able to independently direct the expression of mRFP1 in *C. necator*. However, *E. coli* and *C. necator* transformation efficiencies and colonies expressing reporter protein on agar plate were significantly lower in comparison to that obtained for pCAT201–207 plasmids (with both the *oriV* and *Rep* directing replication). It is noteworthy that the plasmid yields from this set of plasmids (pCAT208–214), with partial replication machinery, were slightly higher and more variable compared to that obtained for plasmids with complete replication machinery (pCAT201–207) (Fig. 4.10). Therefore, the inability of the pBHR1 *Rep* in pCAT101–116 plasmids to direct the expression of mRFP1 in *C. necator* could not be ascribed to the *Rep* being entirely ineffective, since the *Rep* was able to independently direct replication both in the pBHR1 and in pCAT101–116 plasmids. Although it is plausible that the upstream 85 bp of the *Rep* in pCAT101–116 plasmids interfered with regulatory function of these plasmids during replication—resulting in plasmids with high yields from *E. coli* that were unable to transform *C. necator*—this was tested and found not to be entirely the case. This was tested by assembling plasmids (pCAT215 and 216) with only the coding sequence of *Rep*, carrying no truncated *oriV* sequence, to direct the replication of *TcR* with MCS or mRFP1 as the

cargo. Cognisant that coding for tetracycline will exert more metabolic burden in *C. necator* based on the results of antimicrobial susceptibility test (Table 3.4, 3.5 and 3.6).

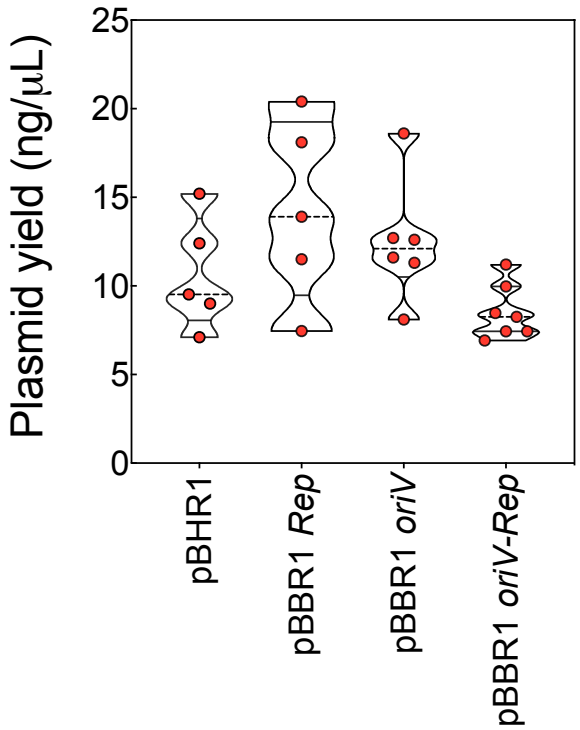


Fig. 4. 10 Effect of replication sequence on plasmid yield.

For each plasmid, 1.5 mL overnight culture was purified and eluted in 50 μL molecular grade water.

The resulting plasmids (pCAT215 and 216) transformed *E. coli* with high efficiency unlike the first assembly counterparts (bearing pBBR1 incomplete replication sequence, *TcR* and MCS or mRFP1 as the cargo), which were unable to transform *E. coli* even after repeated attempts. However, the plasmid yield recovered for pCAT215 and 216 were significantly higher, and subsequent transformation of *C. necator* with these plasmids yielded no transformants (n = 3 biological replicates). This indicates that though the partial upstream *oriV* 85 bp in pCAT101–116 *Rep* affected plasmid propagation in *E. coli*, is not responsible for the unusual high plasmid yields. It is highly probable that the unusual high plasmid yield is due to weak regulation of plasmid copy number under metabolic burden induced either by the expression of reporter protein or antibiotic (tetracycline) resistant. Unfortunately, the outcome of such weak regulation is not well tolerated in the destination host, *C. necator*. Therefore, complete replication system of each plasmid part is crucial for the establishment and controlled propagation of a plasmid, especially in the expression of goi. Subsequently, the remainder of plasmids were assembled using the complete replication sequence (pBBR1 *oriV-Rep*).

4.3.7 Testing the robustness of modular minimal plasmids

Aware that the reporter cassettes (P_{j5} mRP1) and ($P_{j5[C2]}$ mRFP1) were characterised in *C. necator*, the ability of pCAT plasmid system to express other reporter cassettes not previously characterised in *C. necator* was tested. Another set of plasmids with different reporters driven by different promoters (pCAT217–221) was assembled. The promoter P_{j5} was substituted with P_{tac} to obtain $P_{tac[2]}$. Another promoter ($P_{j23100PETRBS}$) extensively characterised in *E. coli* was also tested. The resulting plasmids were delivered to *C. necator* with high transformation efficiency (Table 4.3), and the reporters (mRFP1 and eGFP) well expressed in the bacterium (Fig. 4.11A). The fluorescence output for mRFP1 and eGFP under $P_{j23100PETRBS}$ promoter was significantly higher ($p < 0.05$) compared to the output of the same reporter proteins under $P_{j5[C2]}$ or $P_{tac[C2]}$ promoters (Fig 4.11B).

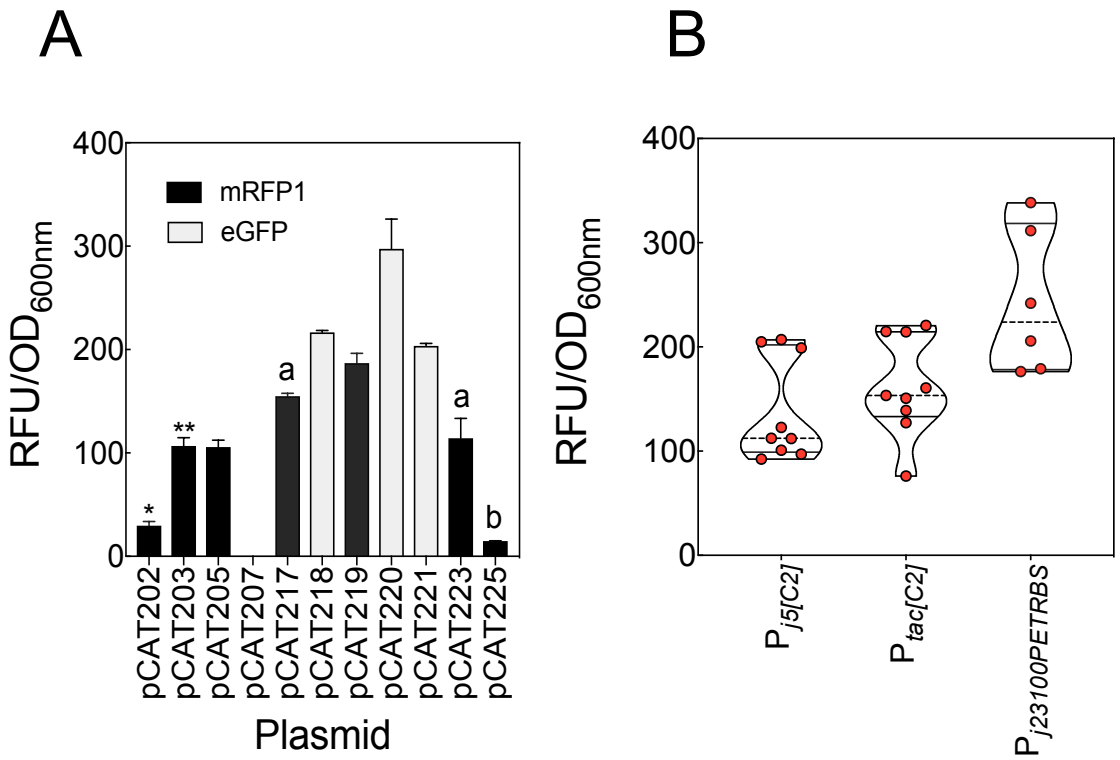


Fig. 4. 11 Characterisation of reporter proteins.

A. Modular minimal plasmids expressing mRFP1 (black filled bars) or eGFP (grey bars). There is a significant difference (**) between non-engineered P_{j5} promoter (pCAT202) and synthetic $P_{j5[C2]}$ counterpart (pCAT203) in the expression of mRFP1. Also, there is a significant reduction (b) in reporter protein expression with plasmid bearing additional origin of replication (pUC19 *ori*, pCAT225); $p < 0.05$. **B.** Combined data for promoter strength in driving the expression of mRFP1 or eGFP in *C. necator*. Error bars are S.E.M., $n = 3$ biological replicates.

Furthermore, the plasmid backbone was extended to four bioparts. In this new system, variants with double antibiotic marker (*KanR* and *CmR*) carrying either MCS or mRFP1 as the cargo (pCAT222 and 223) were assembled. Both plasmids successfully transform *C. necator* and pCAT223 expressed mRFP1 in the bacterium (Table 4.3; Fig. 4.11A).

Additionally, having demonstrated that plasmids with complete replication system is well propagated in *C. necator*, in a final attempt to increase plasmid yield, variants with double replication origin (pUC19 and pBBR1-*oriV-Rep*) were assembled. These variants also carry MCS or mRFP1 as the cargo (pCAT224 and 225, respectively). As expected, the resulting plasmids transformed *C. necator* with high efficiency and expressed the cloned reporter, mRFP1, despite carrying an extra biopart (Table 4.3; Fig. 4.11A). The yield of plasmid recovered from *E. coli* transformed with pCAT224 (bearing MCS as the cargo) was high yet the plasmid successfully transformed *C. necator*. This is not surprising as MCS is likely to exert less metabolic burden compared to reporter protein plus this occurred under compete replication system.

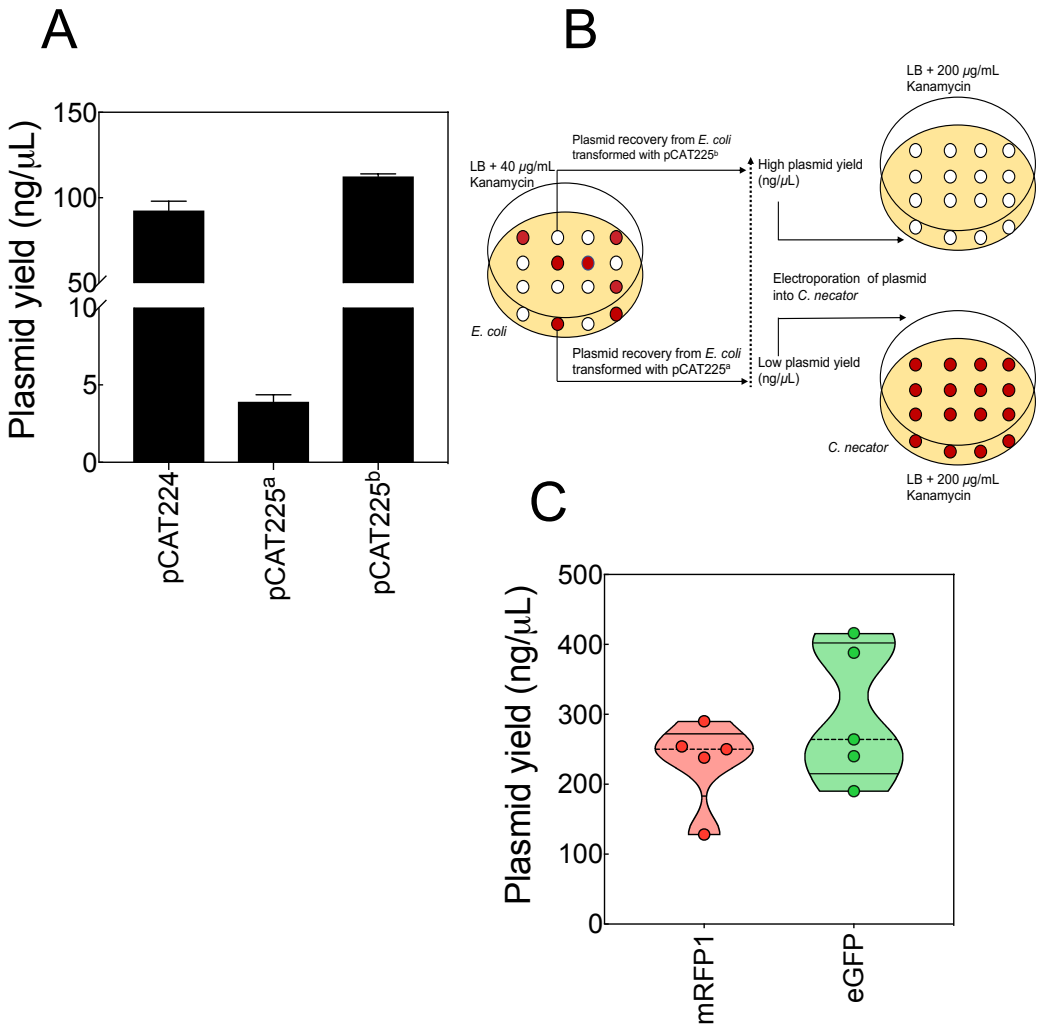


Fig. 4. 12 Effect of high copy *ori* on plasmid yields and gene expression under pBBR1 complete replication sequence.

A. Plasmid yield from variants with complete replication sequence (*oriV-Rep*) and pUC19 *ori*. pCAT224 and pCAT225 bear MCS and mRFP1 as cargo, respectively. Plasmid yield from *E. coli* transformed with pCAT225 and expressing visible mRFP1 on selective LB agar plate is denoted as pCAT225^{a}, whilst that from the same transformation event, on the same agar plate, expressing no visible mRFP1 is denoted as pCAT225^{b}. **B.** Scheme of the effect of pUC19 *ori* on *E. coli* plasmid yield and reporter protein expression in *C. nectaor*. **C.** Combined plasmid yields relative to reporter protein from preliminary assembly. Plasmids carry partial replication sequence (pBHR1 *Rep*) and pUC19 *ori*. Error bars are S.E.M., *n* = 3 biological replicates.

In contrast, the plasmid yields (*n* = 3 biological replicates) recovered from *E. coli* transformants (pCAT225[a]) expressing mRFP1 on agar plate was low, and the yields are like that obtained for pCAT201–221 (variants without pUC19 *ori*). Further, plasmid yields (*n* = 3 biological replicates) recovered from *E. coli* transformants (pCAT225[b]), which showed no evidence of mRFP1 expression on the same agar plate was high—more than 30-fold higher than the yields obtained for transformants expressing mRFP1 on agar plate (Fig. 4.12A). As expected, the high yield pCAT225[b] plasmids were able to transform *C. necator* but did not express the reporter (Fig. 4.12B). This implies that though high plasmid yield is possible with the addition of pUC19 *ori* in pCAT plasmids, the desired transformants that will express foreign gene in *C. necator* have to be under the control of pBBR1 replication system during first replication in *E. coli*, which in turn will result in low plasmid yield. This scenario is favoured under complete (pBBR1 *oriV-Rep*) replication system, which overrides pUC19 *ori* during initial replication in *E. coli* to give low plasmid yield.

During preliminary plasmid assembly with incomplete pBHR1 Rep and pUC19 *ori* (plasmid identity not shown), it was observed that *E. coli* transformants expressing reporters (mRFP1 or eGFP) on agar plates gave very high plasmid yields (Fig. 4.12C). The resulting plasmids were unable to transform *C. necator* after repeated attempts (*n* = 5 Golden gate assembly). Unlike the plasmids from the preliminary assembly, pCAT225[a] recovered from *E. coli* transformants (*n* = 3 biological replicates) expressing mRFP1 on agar plate gave low plasmid yield, which transformed *C. necator* and expressed the reporter protein in the bacterium. Repeated attempts (*n* = 3 biological replicates) to recover plasmids (pCAT225[a]) with high yield from *E. coli* transformants expressing reporter on agar plates were unsuccessful. This observation further complements the earlier hypothesis: the observed high plasmid yield from *E. coli* is more likely to occur under incomplete replication system. Under this condition, metabolic burden induced by the expression of reporter or antibiotic resistant promotes high plasmid yield and the resulting plasmid is unable to transform *C. necator*. Moreover, all variants carrying MCS as the cargo both under complete and incomplete replication system gave low plasmid yield, except those carrying additional high copy ori, pUC19 *ori*. These variants were able to transform *C. necator*, however, with different level of efficiency. Thus, it is evident that having complete replication system (*oriV-Rep*) helps regulate plasmid propagation especially under strong metabolic burden, expression of foreign gene(s) or strong antibiotic (tetracycline) resistant, as observed in this study. Given that addition of pUC19 *ori* did not yield the intended result—high plasmid yield from *E. coli* transformants visibly expressing reporter on agar plate and subsequent transformation and expression of the reporter in *C. necator*—it is recommended that high copy *ori* be excluded from plasmid intended for propagation in *C. necator*.

4.3.8 Segregational stability

Another important feature of a plasmid is the ability to be stably propagated in a host under prolonged cultivation. Therefore, the ability of pCAT plasmids to be stably segregated under sustained cultivation (over several cultivation intervals) in the absence of antibiotics was tested. First, the stability of pBHR1 (with *KanR* and *CmR*) under selective and non-selective conditions was determined. This plasmid lost its relative stability before the fifth interval, 144 h (Fig. 4.13A). Unlike pBHR1, pCAT plasmids under the same cultivation condition maintained relative stability up to the sixth interval (144 h) both under non-selective and selective conditions. Under these conditions, plasmids with complete replication machinery (pBBR1 *oriV-Rep*) carrying *KanR* were stably propagated in *C. necator*, regardless of whether MCS or reporter protein was the cargo (Fig. 4.13B and C). Interestingly, even at the sixth interval (144 h), transformants showed evidence of reporter protein expression both on selective and non-selective agar plates (Fig. 4.14). This further demonstrates the stability of pCAT plasmids bearing pBBR1 *oriV-Rep* and *KanR*.

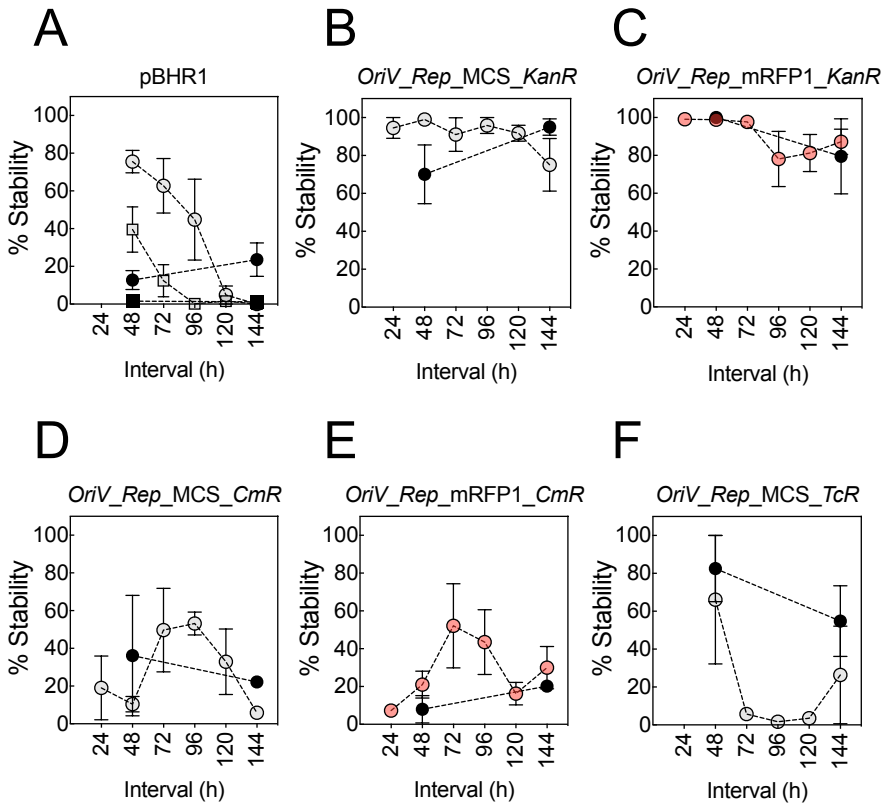


Fig. 4. 13 Effect of antibiotic cassette and reporter protein on plasmid segregational stability in *C. necator*.

A. pBHR1 carries *KanR* and *CmR*. Plasmids cargo and antibiotic cassette as follows: **B.** pCAT201 (MCS and *KanR*); **C.** pCAT203 (mRFP1 and *KanR*); **D.** pCAT204 (MCS and *CmR*); **E.** pCAT205 (mRFP1 and *CmR*); **F.** pCAT 206 (MSC and *TcR*). All plasmids were under complete (*oriV-Rep*) sequence except for pBHR1. Grey or light red circles are for cultivations under non-selective conditions, whilst closed dark circles are for cultivations under selective conditions—with percentage stability determine at the first and sixth interval. For pBHR1, circle represents cultivation with kanamycin as the selective pressure, whilst square is for cultivation with chloramphenicol as the selective pressure. It is noteworthy that pBHR1 and pCAT206 did not grow well after 24 h resulting in 144 h as the fifth interval. Error bars are S.E.M., *n* = 3 biological replicates.

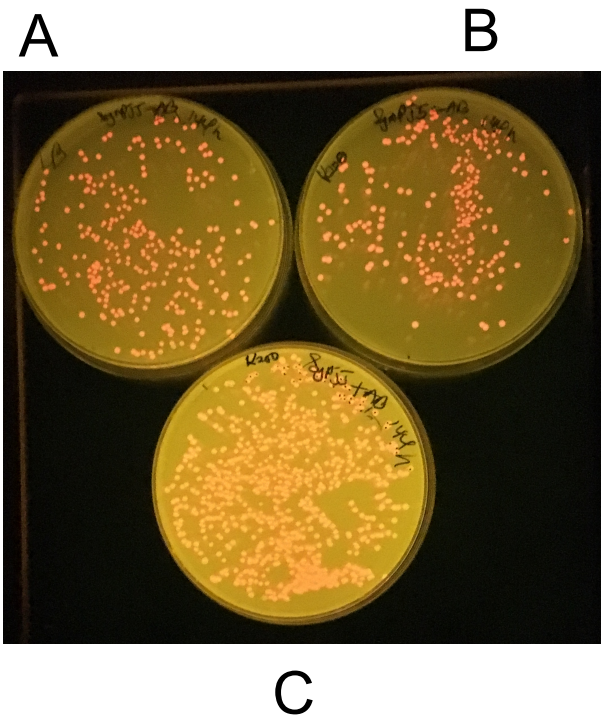


Fig. 4. 14 *C. necator* transformants expressing mRFP1 after six generation.

A. Transformants from culture tube maintained under non-selective condition and spread on LB agar plate supplemented with no antibiotic. **B.** Transformants from tube maintained under non-selective condition and spread on LB-antibiotic agar plate. **C.** Transformant from culture tube maintained under selective condition and plated on LB-antibiotic agar plate.

Conversely, pCAT plasmids with pBBR1 *oriV-Rep* and *CmR* as the resistant cassette showed gradual decrease in segregational stability (Fig. 4.13D and E). By the sixth interval (144 h), the average percentage stability obtained were 5.86 ± 1% and 29.97 ± 11% for variants carrying MCS or reporter protein as the cargo, respectively. Even under selective condition (cultivation in LB-chloramphenicol broth followed by spreading on LB-chloramphenicol agar) there were gradual reductions in segregational stability. The average % stability was 22.22 ± 4% and 20.26 ± 1% for plasmid with MCS or reporter protein as the cargo, respectively (Fig. 4.13D and E). Further, variant bearing pBBR1 *oriV-Rep* and *TcR*, with MCS as the cargo also showed reduction in stability; one of the replicates formed no colony by the sixth interval (144 h), resulting in average % stability of 26.4 ± 25.7% (Fig. 4.13F). This gradual decrease in % stability was also observed under selective cultivation; the stability under this condition was 54.77 ± 34%. It is noteworthy that *C. necator* transformed with plasmids bearing *CmR* were incubated for 72 h to allow transformants to form, unlike the same strain of bacterium transformed with plasmids bearing *KanR* or *TcR*, which formed transformants within 48 h. In addition, it took *TcR* transformant more than 24 h to grow both in selective and non-selective broth.

Furthermore, pCAT plasmids with incomplete replication sequence (pBBR1 *oriV* or pBBR1 *Rep*), bearing the same *KanR* and the same reporter protein showed reduction in % stability under non-selective condition (Fig. 4.15). The average % stability (when mRFP1 is the cargo) were 30.35 ± 3 % and 56.13 ± 13% for plasmid with *oriV* or *Rep* sequence, respectively (Fig. 4.15A and B). Nevertheless, the plasmids retained stability when cultivated under selective conditions. Similar plasmid variant with *Rep* sequence and with MCS as the cargo was stably propagated under non-selective condition for five intervals (Fig. 4.15C). Therefore, it can be inferred that plasmid segregational stability is primarily a function of the resistant cassette and the replication sequence. In addition, with reference to pBHR1, it can be deduced that reduction in plasmid size increased segregation stability, as pCAT212 (Fig. 4.15C) carrying similar incomplete replication (pBBR1 *Rep*) sequence without reporter protein as the cargo was stably propagated for five intervals in the absence of antibiotic as the selection pressure.

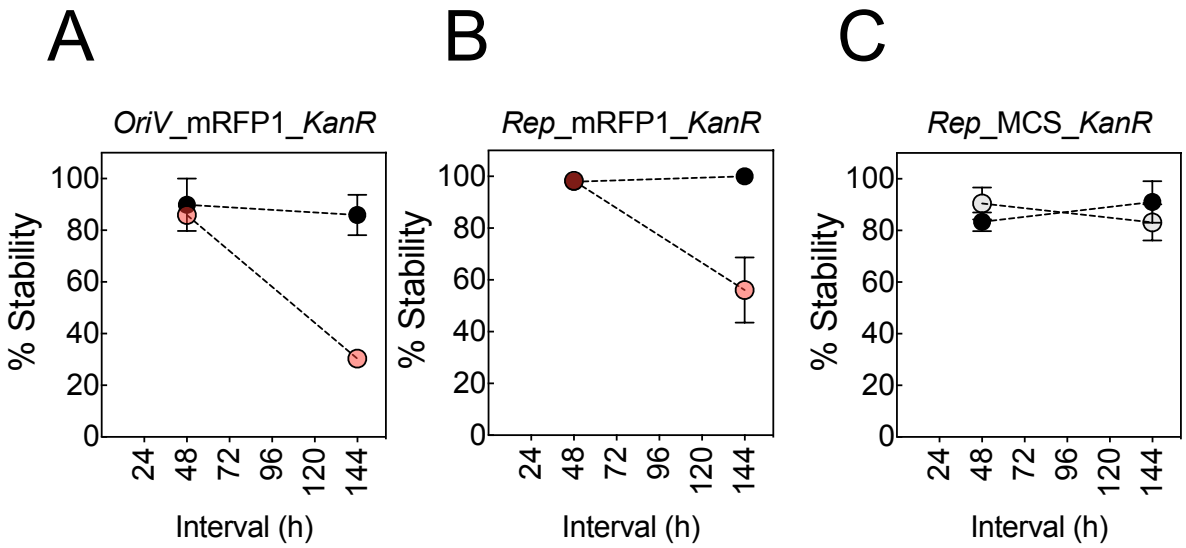


Fig. 4. 15 Effect of partial replication sequence on plasmid segregational stability in *C. necator.*

Plasmids bear replication sequence and cargo as follows: **A.** pCAT209 (*oriV* and mRFP1); **B.** pCAT213 (*Rep* and mRFP1); **C.** pCAT212 (*Rep* and MCS). Transformants expressed mRFP1 throughout the serial cultivation but gradually lost stability over time. pCAT212 retained stability over time. All plasmids carry *KanR*. Stability was determined at the 2nd and sixth interval (48 and 144 h, respectively). Grey or light red circles are for cultivations under non-selective conditions, whilst closed dark circles are for cultivations under selective conditions. Error bars are S.E.M., *n* = 3 biological replicates.

4.3.9 Electroporation of C. necator with digestion-ligation reaction

The objective so far has been to construct modular minimal plasmids that can be delivered to *C. necator* by electroporation with high transformation efficiency. Having achieved this, attention was shifted from delivering assembly reactions via *E. coli*, to delivering them directly to *C. necator*. Earlier in this study, it was observed that *C. necator* transformation efficiency,

under incomplete replication system, from direct electroporation of assembly reaction (pCAT101) was low (4.62 x 10^2 ± 10% transformants/μg DNA). Evident from subsequent results, transformation efficiency obtained for plasmids with pBBR1 *oriV-Rep* is significantly higher in comparison to that obtained for plasmids with partial replication sequence (pBBR1 *oriV* or pBBR1 *Rep*) (Fig. 4.16). With this understanding, pCAT201, 203 and 220 were reassembled and the resulting Golden gate restriction-ligation reaction delivered directly to *C. necator*. Intriguingly, the transformation efficiencies obtained were satisfactory when compared with that obtained for the same plasmid variants delivered as plasmid recovered from *E. coli* (Fig. 4.17A–C). The average transformation efficiencies were 7.59 x 10^4 ± 25%, 6.75 x 10^4 ± 9% and 7.24 x 10^4 ± 16% transformants/μgDNA for pCAT201, 203 and 220 respectively. These efficiencies are more than 11-fold higher compared to that obtained for pBHR1 plasmid recovered from *E. coli*. Moreover, *C. necator* transformed with pCAT203 and 220 Golden gate assembly reactions expressed their individual reporter proteins on agar plates (Fig. 4.17B and C). This efficient direct electroporation of *C. necator* with restriction-ligation reaction offers many advantages. Not least, it saves extra 48 h of cloning in *E. coli* and subsequent plasmid isolation and purification. It also allows direct screening of libraries in *C. necator*.

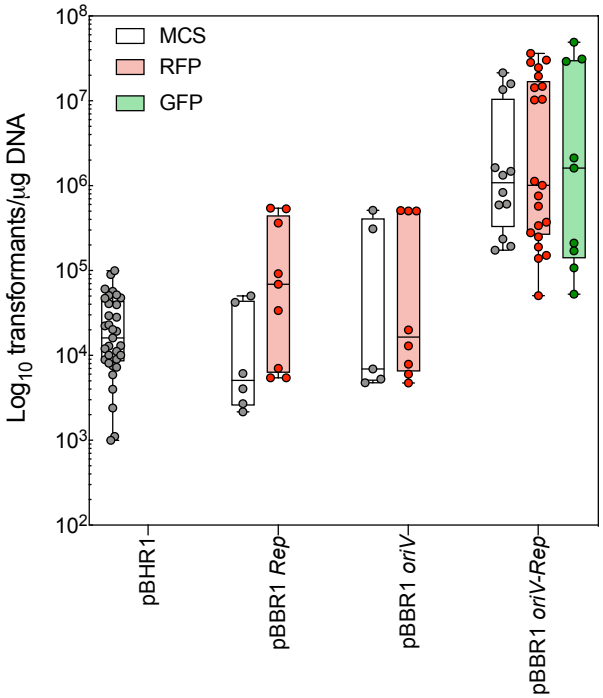


Fig. 4. 16 Comparison of *C. necator* transformation efficiency under different replication sequences.

Furthermore, with the advantage of rapid assembly of plasmids and direct electroporation of *C. necator* with Golden gate assembly reactions, the possibility of assembling a plasmid with more than one functional reporter protein was explored. In line with this, a new plasmid (pBBR1*oriV-Rep*_$P_{j5[C2]}$ mRFP1_$P_{j23100PETRBS}$ eGFP_*KanR*) was assembled and the assembly reaction electroporated directly to *C. necator*. Interestingly, the resulting assembly reaction transformed *C. necator* with high efficiency and both reporters (mRFP1 and eGFP) carried on a single plasmid co-expressed on agar plate. Transformants with an intermediate colour between the two reporters were also observed (Fig. 4.17D). This assembly was designated pCAT226.

Lastly, to validate the ability of pCAT226 transformants to simultaneously express each individual reporter, four mRFP1, four eGFP and ten (mRFP1 and eGFP) intermediate transformants were cultivated. pCAT201, 203 and 220 served as the controls. Following incubation, the control plasmids (pCAT203 and 220) expressed only their individual reporter, with no trace of eGFP detected for pCAT203 and vice versa (Fig. 4.17E). Interestingly, both mRFP1 and eGFP were simultaneously expressed from single transformants resulting from pCAT226 assembly (Fig. 4.17F). Moreover, the expression of mRFP1, eGFP or the intermediate colour between the two reporters was also visible in the cell pellet (Fig. 4.17G and H). Therefore, the objective of building modular minimal plasmids that can be delivered directly to *C. necator* via electroporation and express more than one reporter protein carried on a single plasmid construct was achieved.

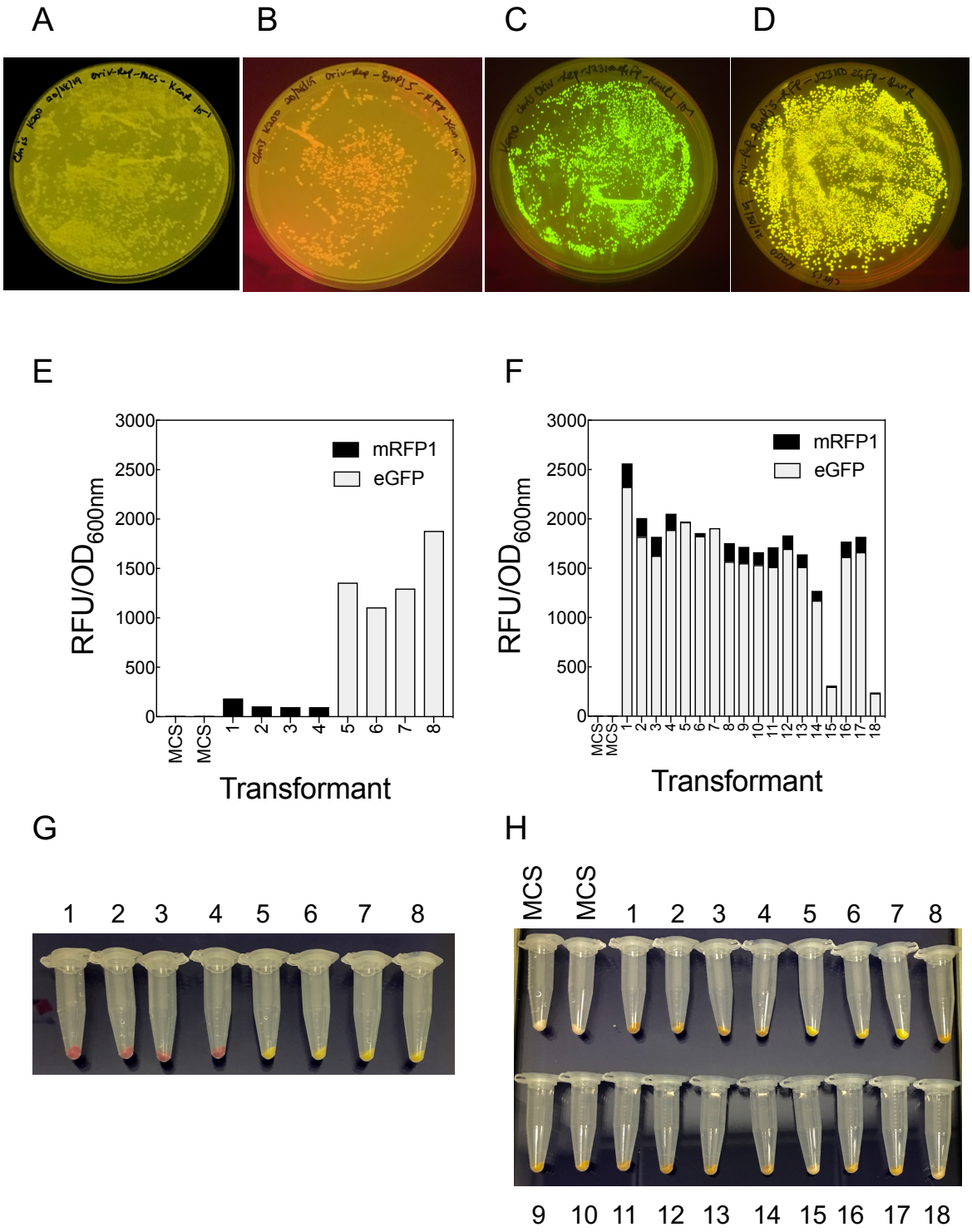


Fig. 4. 17 Direct electroporation of *C. necator* with Golden gate assembly reaction.

A. pCAT201. **B.** pCAT203. **C.** pCAT220. **D.** pCAT226. **E.** Fluorescence output of four pCAT203 transformants (1–4) and four pCAT220 transformants (5–8) after 48 h. **F.** Co-expression of mRFP1 and eGFP from single pCAT226 transformants. A total of 18 transformants were picked from agar plate D and cultivated for 48 h. From plate D: transformants expressing mRFP1 on agar plate (1–4); transformants expressing eGFP (5–8); transformants expressing intermediate colours for mRFP1 and eGFP (9–18). **G.** Bacterial pellet of transformants expressing either mRFP1 (pCAT203) or eGFP (pCAT220); samples tubes are as described for E. **H.** Bacterial pellet of pCAT226 transformants co-expressing mRFP1 and eGFP; sample tubes are as described for F**.** For each tube, 3 mL of 48 h culture was harvested. pCAT201 transformants, with MCS as the cargo, served as the negative control.

4.4 Discussion

This study has systematically characterised and standardised plasmids for *C. necator*, resulting in novel modular broad host range minimal plasmids with improved transformation efficiency, segregational stability and expression of reporter proteins in the bacterium. Standardisation of plasmid bioparts for this microbial chassis, *C. necator*, offers the opportunity for high-throughput construction and de-/re-construction of genetic construct(s) intended for downstream applications in the bacterium. It further reduces the complexity and uncertainty encountered whilst working with non-standardised plasmid vectors. The standardisation together with the high-throughput cloning method, Golden gate, [161] offers more flexibility. Each biopart can be used for another microbial host independent of pCAT plasmid backbone. For example, using the same minimal plasmid backbone, plasmids with other antibiotic resistant markers (ampicillin, gentamicin or erythromycin) can be assembled for suitable hosts that are not resistance to these antibiotics. One of the existing plasmids (pBBR1MCS-2) with the same broad host range replicon as that of pCAT plasmids was tested and found to replicate in different bacterial genus: *Acetobacter*, *Bartonella*, *Bordetella*, *Brucella*, *Caulobacter*, *Escherichia*, *Paracoccus*, *Pseudomonas*, *Rhizobium*, *Rhodobacter*, *Salmonella*, *Vibrio*, and *Xanthomonas* [48]. Therefore, pCAT plasmids can potentially be propagated in different Gram-negative microbial hosts and can be delivered by electroporation and possibly heat-shock.

Plasmids that were recovered with high yields from *E. coli* impaired *C. necator* transformation efficiency, especially variants carrying reporter proteins. Similar observation have been reported previously [51]. This unusual high plasmid yields were mostly observed when reporter proteins were the cargo. The unusual high plasmid yield with pBBR1 origin of replication—medium copy number replicon—is attributed to the metabolic burden induced by the expression of the reporter proteins under incomplete replication system. Some of the observed effects in *C. necator* are low or no transformation efficiency, and suboptimal growth and expression level at 24 h. Additionally, the ration of *C. necator* transformants (in the case of pCAT208, 209, 212, and 213) expressing visible reporter on agar plates to overall transformants on selective agar plates was low (< 40%), arguing that expression of a foreign gene in the bacterium might have elicited metabolic burden, which the bacterium counteracted by reallocating resources to survival (colony formation) rather and gene expression.

Similar metabolic burden in *E. coli* elicited physiological response, which exacerbates under incomplete replication machinery, particularly under condition when *Rep* is directing

replication. Unlike *C. necator*, the observed effect of similar metabolic burden in *E. coli* was mostly high plasmid yield; there was no considerable difference in growth and gene expression level at 24 h. Unfortunately, the modifying effect of such response on the plasmids could not be tolerated in the destination host, *C. necator*. One of the possible modifying effects in *E. coli* is mutation [176]. High copy number mutant plasmids in *E. coli* could not be established in other hosts, *Pseudomonas aeruginosa* and *Azotobacter vinelandii*, due to bacteria having different upper tolerance level to the side effect of high copy number [176]. The notable serious side effect was cell death. The high copy number occurred due to physiological response and reduced growth rate under the conditions that induced the mutation. High intracellular concentration of a trans-acting replication function (*TrfA* or *Rep*) was implicated to enhance such mutation. This agrees with the observation that the unusual high plasmid yield was obtained under the condition when *Rep* is solely directing replication. Given that bacteria may employ the same mechanism of copy number regulation, it follows that *E. coli* may have upper tolerance level to the side effect of such mutation compared to *C. necator*. Moreover, plasmid topology varies across bacterial hosts [47]. Another possible explanation is that the plasmid may be employing an iterons—short base pair repeats within the origin of replication (*oriV*) or auxiliary iterons located outside the *oriV*—to control plasmid replication [177,178]. Under high plasmid copy number, which brings about high concentration of iterons, auxiliary iterons acts as a negative feedback regulator of plasmid copy number by turning off plasmid replication [178]. This is highly plausible as no unusual high plasmid yields were obtained when the replication system contained both the *Rep* and the *oriV*. Rather, plasmid yields were low, leading to high transformation efficiency, expression of reporter protein and stable segregation in *C. necator*. Plasmid with improper origin of replication for a bacterial host, tends to lose stability and ability to autonomously replicate the genetic material it encodes [132]. More so, circuit implemented in low copy number displayed better stability and minimal metabolic burden in a microbial host [179].

The inability of plasmids with pMB1 (pBR322 and pUC19) and p15A *ori* to direct replication in *C. necator* is attributed to the narrow host range nature of the *ori* [180]. Replication of plasmids bearing such *ori* is restricted to *E. coli* and other related enterobacterial spp. This is because these *ori* are mostly recognised by the *E. coli* replication protein [47,132,181]. Unlike narrow host range plasmids, broad host range plasmids encode their own replication initiation protein (Rep), which recognises their *ori* within a plasmid thus allowing autonomous replication across different hosts. Plasmid pBBR1 was originally reported to contain two functional cassettes: the replication and the mobilisation regions [48,49]. The replication region contains sequences for both the *oriV* and the *Rep*, which are located together. Plasmid *Rep* is believed to be essential for replication, especially for broad host range plasmid [182]. Yet, this

study showed that either the *Rep* or the *oriV* of pBBR1 can independently direct replication across Gram-negative bacteria. Although pBBR1 is used across Gram-negative bacterial chassis, its mechanism of replication is not fully understood, and no incompatibility group has been assigned to the plasmid. Further genetic studies are required to gain more insight into the plasmid behaviour in a host; this will be very useful in future applications of the plasmid. It is possible that the so-called *Rep* and *oriV* are incorrectly mapped or annotated. Additionally, it is possible that they are neither separate *oriV* nor *Rep*, rather a single genetic module coding for the double stranded origin (dso)—the *Rep* gene and the elements involve in plasmid control [177,183]. However, this will only hold true if the plasmid is established to be replicating by the rolling circle mechanism of replication.

Few studies have focused on optimising the transformation efficiency of *C. necator* by electroporation [56,73,74]. Optimal electroporation conditions were established to be 0.8 cell density (OD_{600nm}) at the time of harvest, 12.5 kV/cm field strength, 200 Ω, 25 μF and a single pulse time of ~ 5 ms [74]. Treatments such as transformation buffer containing 0.2 M fructose, and incubation of cells in 50 mM $CaCl_2$ for 15 min during preparation of electrocompetent cells improve transformation efficiency [73]. Both studies showed that higher concentration of DNA (1 *μ*g) is required for higher transformation efficiency. This is in contrast to the findings of this study, which observed that lower DNA concentration (30–100 ng) is sufficient to obtain higher transformation efficiency. Further, some restriction endonuclease genes were identified and deleted in *C. necator* resulting in improved transformation efficiency [56]. Nevertheless, the transformation efficiency obtained in this study was 42 and 338-fold higher in comparison to best transformation efficiencies obtained by [73] and [56], respectively. This study showed that compatibility of plasmid bioparts (complete replication system and suitable antibiotic resistance cassettes) are key to obtaining high transformation efficiencies.

As expected, variants of plasmids with *KanR* showed higher transformation efficiencies and better stability in comparison to variants with *CmR* or *TcR* as the resistant marker. The differences in transformation efficiencies and segregational stability were attributed to the antibiotics rather than the individual cassette. Amongst the three antibiotics, *C. necator* wild type was most sensitive to tetracycline followed by chloramphenicol and then kanamycin. Thus, it is likely that coding for antibiotics that the bacterium is very sensitive to exerted more metabolic burden on the bacterium, despite being transformed with plasmid coding for resistance against such antibiotics. This implies that the search for antibiotics with very high antimicrobial activity against a microbial chassis, with the objective of selecting transformants at low antibiotic concentration, may not translate to higher plasmid establishment and maintenance in a given microbial host.

4.5 Conclusion

Modular minimal plasmids bearing broad host range replication system were designed and assembled in a high-throughput approach. The resulting plasmids were characterised for transformation efficiency, segregational stability and expression of reporter genes, in a model chemlithoautotrophic bacterium, *C. necator*. pCAT plasmids will expand the genetic tools available for genetic manipulation of *C. necator* and will resolve the challenges: low transformation efficiency, ineffective expression of foreign gene(s) and poor segregational stability associated with cloning in the bacterium. Each biopart of pCAT plasmids was extensively characterised, thus facilitating a reliable and predictable outcome during downstream application. The modularity of pCAT plasmids together with the assembly method would allow future high-throughput plasmid construction using pCAT as the backbone vector. Further, pCAT bioparts can be used as suitable bioparts for refactoring other plasmid vectors. In chapter three, heat-shock transformation of *C. necator* was demonstrated and found plasmid backbone, specifically yield, as one of the factors impeding further improvement in transformation efficiency. This is based on the observation that plasmids with high yields are not well established in the bacterium and that the pBBR1 replication system is a medium copy number, which gives a low plasmid yield after purification. Therefore, it is highly improbable to obtain the required plasmid concentration following plasmid isolation and purification without further treatment. Hence, optimising the protocol would likely be impeded by currently existing plasmid backbone. Nevertheless, purified assembly reaction (restriction-ligation mixture) can be delivered directly to *C. necator* by electroporation (without passing through *E. coli*) to achieve high transformation efficiency. As demonstrated in this study: the ineffective transformation of *C. necator* with pBBR1MCS-2 is due to the kanamycin cassette in the plasmid; the low transformation efficiency associated with pBHR1 is due to the plasmid incomplete replication system; reduction in plasmid size improved transformation efficiency; plasmids with complete replication system are better maintained in *C. necator* over several generations of cultivation. The plasmid resource from this study (Table 4.3) has been sequence verified and is freely accessible under the Addgene Repository https://www.addgene.org. The toolkit reported in this study are key to advancing bioengineering applications in *C. necator*, an industrially relevant lithoautotrophic microbe. Thus, allowing its deployment as a microbial chassis for many industrial applications.

Chapter 5.

Global Transcription Machinery Engineering of *Cupriavidus necator* H16 for Improved Tolerance to Biofuels

5.1 Introduction

Interest in biofuels as an alternative source of energy over petroleum-derived energy source is gaining more attention both in developing and developed countries [184–186]. This is driven by government, economic and environmental policies aimed at reducing pollution and environmental degradation caused predominantly by the use of petroleum-derived fuels. Unlike petroleum-derived energy, the use of biofuel results in almost zero emission of greenhouse gases [187]. Despite the environmental benefits of biofuels, first-generation biofuels are not economically competitive with fossil fuels due to limited resources such as feedstock and land [187,188]. Second-generation biofuels overcome some of the major limitations of first-generation biofuels by utilising plant biomass (lignocellulosic material), which is cheap and abundant. However, production of such biofuels still faces some technical barriers (cost of feedstock production, pre-treatment technologies and biological conversion), which in turn renders the product less competitive with fossil fuels [188–190]. Third-generation biofuels overcome the limitations associated with both the first- and second-generation biofuels. Microalgae are considered good candidate for third-generation biofuels owing to their high lipid content, as well as their fast growth rate [191]. However, producing microalgal biomass is less economically favourable, requiring special growth conditions that must be met and sustained for high turnover [189]. Moreover, biofuels are not completely compatible with most combustion engines [192]. To meet the compatibility requirement, pathways for the synthesis of biofuel, with similar chemical and structural moiety as with petroleum-derived fuels, need to be conceived, designed and assembled in a microbial chassis. This approach has been demonstrated in *Escherichia coli* for the synthesis of saturated and unsaturated hydrocarbons [193–196]. Despite recent advances in biotechnology to produce customised biofuels, toxicity of biofuels to production host is a major limitation affecting productivity [197–199]. One of the possible ways of overcoming this challenge is to engineer the production host, at the cellular level, to be tolerant to the target biofuel. To this end, several engineering approaches including mutation have been employed to generate microbial strains (yeasts and bacteria) with desired phenotype that is beneficial for biofuel synthesis [200–203].

The procedure for generating microbial host with desired phenotypic trait for biosynthesis is often tedious, and in some cases the optimum phenotypic trait is not globally attained owing to the intricacy of microbial metabolic landscape [204]. In bacteria DNA-directed RNA polymerase sigma subunit (*rpoD*) gene, coding for the sigma factor σ^{70}, and other sigma factors are often targeted by error-prone polymerase chain reaction (PCR) random mutagenesis to achieve mutation that can regulate transcriptome at global level, hence, global transcription machinery engineering (gTME) [201]. This approach relies on refocusing promoter preference of RNA polymerase during transcription by inducing mutations in sigma factors. Such mutation(s) has generated strains with varying degree of tolerance (phenotypes) to different stress including pH [204], oxidative stress [205,206], biofuel [201,202] and other solvents [203,207,208]. This approach, gTME, has been extended to different microbes other than yeast and *E. coli* to engineer strains with desired phenotype [209].

In *Saccharomyces cerevisaie*, random mutagenesis of transcription factor *SPT15* (a TATA-binding protein) and one of the associated factors (*TAF25*) selectively resulted in improved tolerance of the microbe to elevated concentration of ethanol (> 15 % v/v) and glucose (120 g/L), without alteration in other chromosomal genes [210]. The amino acid mutations responsible for the desirable phenotype were located within repeat element and were: $Phe^{177}Ser$, $Tyr^{195}His$ and $Lys^{128}Arg$. Similarly, in *E. coli*, gTME approach demonstrated a more rapid way of unlocking complex desirable phenotype in bacteria compared to conventional methods [201]. Several rounds of gTME in *E. coli* σ^{70} resulted in superior production strains with more than one desirable phenotype due to multiple amino acid mutations. The resulting strains, unlike the wild type, survived in 70 g/L ethanol and elicited improved lycopene production.

C. necator H16 has two *rpoD1* genes, both located on chromosome 1 (GenBank: AM260479.1). The amino acid sequences for both genes differ significantly (Table 5.1). The shorter of the rpoD (locus tag H16_A1626) with protein reference Q0KB63 has 656 amino acid sequence, whilst the other H16_A2725 (Q0K867) has 827 amino acids (Appendix D). H16_A1626 is flanked by a predicted glutamine amidotransferase and a hypothetical membrane associated protein, while H16_A2725 is flanked by a DNA primase (*dnaG*) and a phospholipase C (*plcN1*). It is noteworthy that the DNA start codon for *dnaG* and one of the rpoD1 genes, Q0KB867, are non-canonical valine (V, encoded with GTG). The implication of the non-canonical start codon, and whether both rpoD are functional in the bacterium are yet to be established. Application of *C. necator* is extending beyond biopolymer production to solvent production, including but not limited to biofuels [8,23]. The bacterium has been engineered to produce alkanes [50] and alcohols [23]. However, as with other microbes, toxicity

of target compound to *C. necator* is a limiting factor affecting productivity. Therefore, the aim of this study is to engineer *C. necator* for improved tolerance to biofuel molecules using gTME approach. This provides great opportunity for producing high titre of biofuels and potentially more complex biofuel molecules from *C. necator*.

Previously, it was demonstrated that plasmids can be rapidly assembled in a high parallel fashion and the resulting Golden gate restriction-ligation products (plasmids) can be delivered directly to *C. necator* with high transformation efficiency Chapter 4. This provides an opportunity for prototyping new phenotypes in *C. necator*—overcoming the limitations of plasmid construction and low transformation efficiencies, which preclude efficient metabolic engineering approaches in the bacterium. Using the tools established in Chapter 3 and 4, *C. necator* σ^{70} (*rpoD1*) were mutated using error-prone random mutagenesis PCR and screened for improved tolerance to range of alcohols and hydrocarbons.

Table 5. 1 Multiple sequence alignment of two *rpoD* genes present in *C. necator* H16

```
Q0KB63      ------------------------------------------------------------    0
Q0K867      MKTTAAKTSTASVKTPARSAPSKSVRSAQAPAVTTPARTRTSEKTKASTSARESGKSAST    60

Q0KB63      ------------------------------------------------------------    0
Q0K867      SGKVAASAPVKGKSITVAKQQNTEVESKRAAAAGAKSGEAKAGTTTTATKARAATPASVA    120

Q0KB63      ---------------------------------MTSGV-----------------QR---    7
Q0K867      SERPAAPAIKPEPKKRGRKPKAEMQHDDSTTDDVTEEFYENDARPAATPAAAPKTEKQKA    180
                                             :*. .                 ::

Q0KB63      TAAKTRETGAIKDPGQTSAGSTAQPDPSTEAAARSQQLRALIQLGRQRGYLTHADISDHL    67
Q0K867      KDRKAKEKALLKEF------ASTQQGTEEELELRRQKLKALIKLGKSRGYLTYAEINDHL    234
            .  *::*.. :*:       :::* . . *   * *:*:***:**:.*****:*:*.***

Q0KB63      PENFTDTAAMESIVSTFAEMGVKIYEQTPDAETLLLSDG-PVVASDDQADEEAEVALATV    126
Q0K867      PDDMVDSETIDTLVATLNDIGIAVYEQAPDAETLLLNDNAPSATSEEEAEEEAEAALSTV    294
            *::::.*: ::::::*:*: ::*: :***:********.*. * .:*:::*:****.**:**

Q0KB63      DSEFGRTTDPVRMYMREMSSATLLTRKQEVEIAKRIEEGLNNMVHAISACPFTIAAILEL    186
Q0K867      DSEFGRTTDPVRMYMREMGTVELLTREGEIEIAKRIEAGLKDMVMAISACPVTISEILAH    354
            ******************.:. ****: *:******* **::** ******.**: **

Q0KB63      SGKVASNEISIDDLVDGLSDESIAEAAVAAA---ADETDDSVDSADEADDDSEDSDDEGS    243
Q0K867      AERVANDEIKIDEFVDGLIDPNADEAPEAPAAPAAAADDEDIESDDEEEGDEDDDDEGGA    414
            : :**.:**.**::**** * .  **  * *   *   *:.::* ** :.*.:*.*: *:

Q0KB63      TQQSNEKALAQLREECLKRFARVTAQFELMCQESAANGAGSAAFLAARDAVREELRTIRF    303
Q0K867      GAGASARQLEELKQNALEKFRVIAEQFDKMRRAFEKEGYNSKPYVKAQEAIQAELMGIRF    474
               :. : * :*::::.*::*  :: **: * :    :* .*  :: *::*:: **  ***

Q0KB63      TAKTIERLCANVQAMVDEVRTVERQVVQLLVERCGMEREEVIARFPGNETNLAWGQELVA    363
Q0K867      TARNVERLCDTLRGQVDEVRKLERSILNIVVDKCGMPRSEFVARFPGNETNLEWVHTIVA    534
            **:.:**** .::. *****.:**.:::::*::*** *.*.:********** * : :**

Q0KB63      QSRPYSAAVARALPDLEAHQQKLIDIQARAALSLPDLKGVNRKMLAAERQMRQAKHEMTQ    423
Q0K867      DGKGYSTIVERNVPAVHELQQKLIDLQSRVVLPLKELKGVNRKMAEGERRAREAKREMTE    594
            :.: **: * * :* :.  ******:*:*..* * :********  .**: *:**:***:

Q0KB63      ANLRLVISIAKKYTNRGMLFLDLIQEGNIGLMKAVDKFEYRRGWKFSTYATWWVRQAVTR    483
Q0K867      ANLRLVISIAKKYTNRGLQFLDLIQEGNIGLMKAVDKFEYRRGYKFSTYATWWIRQAITR    654
            *****************: ************************:*********:***:**

Q0KB63      AIADQARTIRVPVHMIEQINKLNRLSREIMQQTGKEPDPAVLAERLDMTEDKVRSIMKIA    543
Q0K867      SIADQARTIRIPVHMIETINKMNRISRQILQETGNEPDPATLAEKMEMPEDKIRKIMKIA    714
            :*********:****** ***:**:**:*:*:**:*****.***:::* ***:*.*****

Q0KB63      KEPVSMETPVGEDGDTSLGDMIADSDTATPADAALQAGLRAVVREMLDELTPREAKVLRM    603
Q0K867      KEPISMETPIGDDDDSHLGDFIEDTNTLAPAEAALHGSMRDVVKDVLDSLTPREAKVLRM    774
            ***:*****:*:*.*: ***:* *:::* :**:***:..:* **:::**.***********

Q0KB63      RFGIDMSTDYTLEEVGKQFDVTRERIRQIESKAMKKLRHPSRADQLITYLRDA  656
Q0K867      RFGIEMSTDHTLEEVGKQFDVTRERIRQIEAKALRKLRHPSRSDKLKSFLEGN  827
            ****:****:********************:**::*******:*:* ::*..
```

H16_A1626 CDS (2281576..2283546); H16_A2725 CDS (1108133..1110616). DNA binding region highlighted yellow. CLUSTAL O(1.2.4) multiple sequence alignment of both rpoD protein sequences.

5.2 Materials and Methods

5.2.1 Chemicals, media and molecular reagents

Analytical grade ethanol and isopropanol were obtained from Fisher chemical, whilst isobutanol, 3-methylbutanol, heptane and pentadecane were obtained from Sigma-Aldrich. *C. necator* transformants were selected on lysogeny agar plates supplemented with 200 *μ*g/mL kanamycin. All cultivations were carried out under aerobic condition at 30°C and 200 rpm. Optical density (OD_{600nm}) was measured using a Varioskan LUX™ Multimode Microplate reader (Thermo Scientific). Recovery broth, super optimum broth (SOB) is as described previously and contained 20 mM fructose instead of glucose [Chapter 3.2.1]. *C. necator* chemically defined medium is as described in [Chapter 4.2.1]. Primers were synthesised by Life Technologies. Polymerase chain reaction, fragment purification, DpnI digestion and Golden gate assembly were performed as described in [Chapter 4.2.2, 4.2.3].

5.2.2 Plasmid construction

Genomic DNA was extracted from *C. necator* using GenElute Genomic DNA extraction Kit (Sigma-Aldrich). Plasmid construction was based on pCAT backbone with complete replication sequence (*OriV-Rep*) and *KanR* as the selection marker (Fig 5.2). Primers for amplifying these two bioparts are as described in Table 4.1. Gene *H16_A1626* was amplified with F: 5'-GATGCTCCTCCTTGACCGGT-3' and R: 5'-CCGAAAGCTCCTGTCTGCTC-3', whilst *H16_A2725* was amplified with F: 5'-CTTGGCTGAGGTGCCTGTG-3' and R: 5'-ATTGAGCGGTAGTTCGAGGC-3'. Full sequence of each *rpoD* gene is provided in Appendix D. *C. necator rpoD1* genes, *H16_A1626* and *H16_A2725*, together with their individual native promoters and intergenic regions were amplified from purified genomic DNA using high fidelity proofreading enzyme (Q5 DNA polymerase). Fragments were treated with DpnI and purified using Qiagen purification kit. Next, purified DNA fragments were subjected to random mutagenesis using a GenemorphII Random Mutagenesis kit (Agilent Technologies). Template concentrations varied according to supplier's instruction and they were: low (0–4.5 mutation/kb); medium (4.5–9 mutation/kb) and high (9–16 mutation/kb). Error-prone PCR condition was set at 95°C, 2 min; 95°C, 30 s; 68°C, 30 s; 72°C, 2.5 min (for *H16_A1626* and its intergenic region or 3 min for *H16_A2725* and its intergenic region) and final extension at 72°C, 10 min. Following amplification, fragments were purified and quantified using a Qubit 2.0 Fluorometer (Invitrogen™).

Plasmids were assembled using Golden gate assembly method [161] as described in [Chapter 4.2.3]. Briefly, ~40 ng each rpoD fragment was combined with ~ 60 ng each *oriV-Rep* and *KanR*.

The reaction was allowed to proceed for 30 cycles (37°C, 5 min; 16°C, 10 min) followed by 50°C for 5 min and 80°C for 5 min. Subsequently, reaction products were purified and ~ 30 ng delivered directly to *C. necator* by electroporation. Transformants were selected on LB agar plates supplemented with 200 μg/mL kanamycin after 48 h incubation at 30°C.

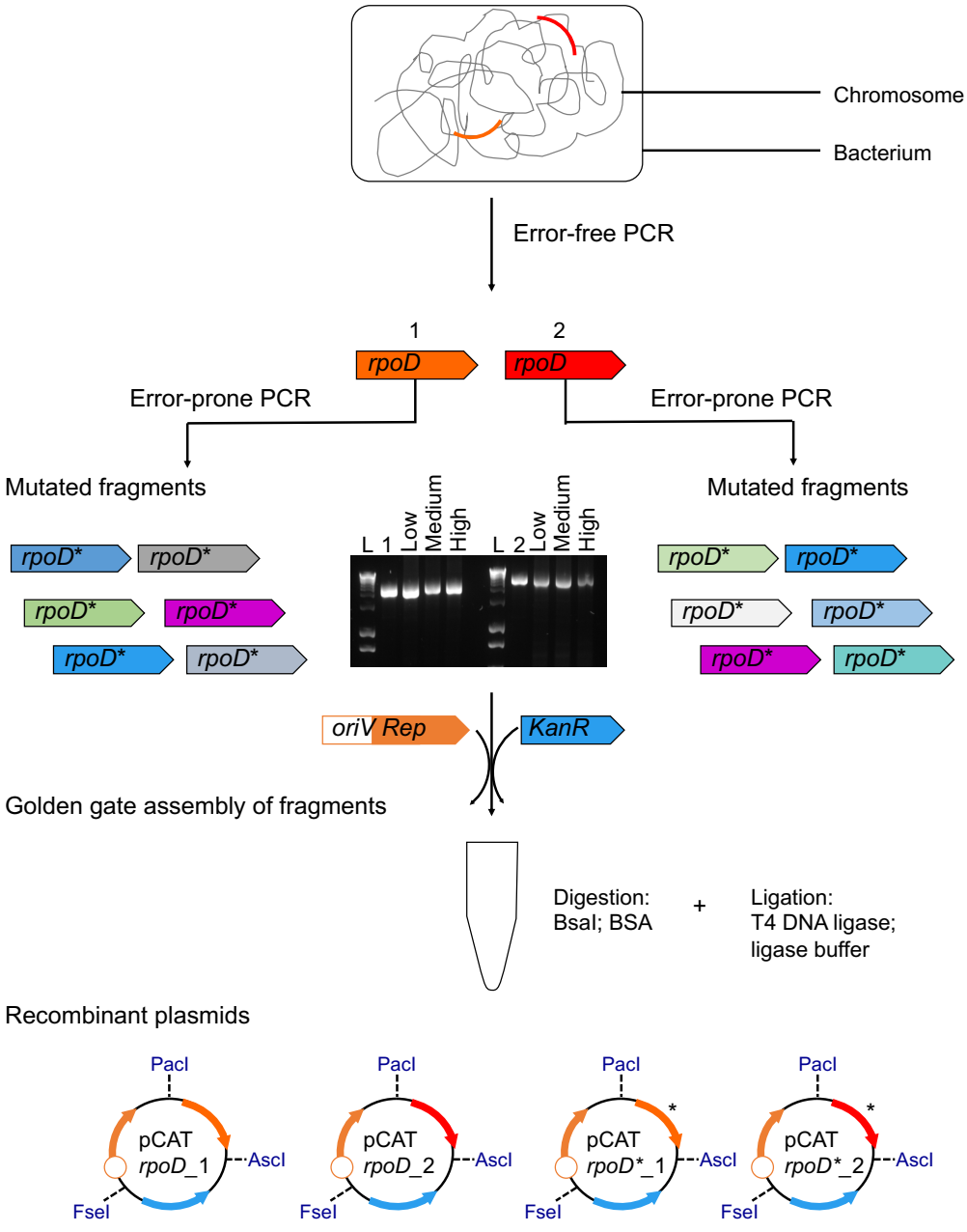


Fig. 5. 1 Construction of *rpoD* recombinant plasmids.

Each *rpoD* together with its intergenic regions was amplified from genomic DNA using a high-fidelity DNA (Q5) polymerase. Fragments were purified and digested with DpnI and subsequently subjected to fragment error-prone PCR, using GeneMorphII random mutagenesis kit. Fragments were analysed on 0.8% gel. Lanes: ladder (L); 1, error-free *rpoD* (Q0KB63 (*H16_A1626*) fragment; 2, error-free *rpoD* Q0K867 (*H16_A2725*) fragment; low, medium and high are mutation frequency range according to supplier's instruction. Fragments were further purified and assembled using golden gate approach. *rpoD* originating from random mutagenesis are detonated with asterisk (*).

5.2.3 Screening for desired phenotypic traits

Phenotype screening was carried out as described [201,202,204]. Briefly, transformants resulting from the three mutation rates were scrapped off into 50 mL falcon tube containing LB broth to create a liquid library. After pooling each library for the individual *rpoD*, glycerol stocks were made and stored at -80°C, while a fraction of the sample was subjected to a challenging environment to select for desired phenotype. To achieve this, 1% of the pooled library was inoculated into fresh LB broth supplemented with 200 μg/mL kanamycin and each of the chemicals tested. After 48 h, depending on level of growth, samples were transferred to fresh LB-kanamycin broth containing higher concentration of chemicals. Also, at this interval glycerol stocks of each culture were made and stored at -80°C for further analysis.

Library resulting from *C. necator* transformed with plasmid containing unmutant *rpoD* served as the control. Additionally, library from plasmid with multiple cloning site (MCS) as the cargo, instead of *rpoD*, was used as the reference to distinguish between the effect of heterologous expression of mutated or unmutated *rpoD* from that of chromosomal version of the gene.

5.2.4 Plasmid isolation from Cupriavidus necator

Isolation of recombinant plasmid from *C. necator* transformants was carried out using Qiagen Minprep kit, with slight modifications to the protocol as described [53,56,211]. One millilitre of overnight culture of *C. necator* transformants was harvested at room temperature. Cells were suspended in resuspension buffer containing RNaseA and incubated for 5 min. Alkaline lysis was performed for 15 min followed by neutralisation for 15 min. Suspension was spun for 20 min and supernatant transferred to spin column and spun for 2 min. The first wash was carried out with PB buffer after 2 min incubation, whilst the second and final wash was carried out with PE buffer containing 80% ethanol after incubating column for 5 min. Plasmids were eluted using 50 μL room temperature molecular grade (Just) water. All centrifugations were performed at 17000 x g and incubation done at room temperature. Plasmids were analysed on 0.8% agarose gel and quantified using Qubit 2.0 Fluorometer with broad host range reagent. Confirmation of mutation in *rpoD* sequence was carried out using the primers shown in Table 5.2.

5.2.5 Sequence analysis

Table 5. 2 Primer for confirming mutation in *rpoD.*

Q0KB63	Q0K867
First design	
F1: 5'- GATGCTCCTCCTTGACCGGT-3'	F1: 5'-CTTGGCTGAGGTGCCTGTG-3'
F2: 5'-CTCCTGAGCGATGGCCCGGT-3'	F2: 5'-CTGCATCCGTTGCATCCGAG-3'
F3: 5'-GTCGACGAGGTGCGCACGGT-3'	F3: 5'-GCCTGAAGGACATGGTGATG-3'
F4: 5'-CTCAATCGCCTGTCGCGCGA-3'	F4: 5'-TCCCGGGCAACGAGACCAAC-3'
R1: 5'-ACCGGTCAAGGAGGAGCATC-3'	F5: 5'-ACAAGATCCGCAAGATCATG-3'
	R1: 5'-CACAGGCACCTCAGCCAAG-3'
Second design	
F1: 5'-GTAAGTGCGCTGTTCCAGAC-3'	F1: 5'-GTAAGTGCGCTGTTCCAGAC-3'
R1:5'-AGCGTTGCCGAGCTCATTTC-3'	R1: 5'-CTCTTCAGTCACGTCGTCTG-3'
F2: 5'-GAAATGAGCTCGGCAACGCT-3'	F2: 5'-CAGACGACGTGACTGAAGAG-3'
R2: 5'-GCCTGGATGTCGATCAGCTT-3'	R2: 5'-CGTCGACGAATTCGTCGATC-3'
F3: 5'-AAGCTGATCGACATCCAGGC-3'	F3: 5'-GATCGACGAATTCGTCGACG-3'
R3: 5'-GGTTGTAACACTGGCAGAGC-3'	R4: 5'-GGCGTAGGTCGAGAACTTGT-3'
	F4: 5'-ACAAGTTCTCGACCTACGCC-3'
	R5: 5'-GGTTGTAACACTGGCAGAGC-3'

5.3 Results

5.3.1 Tolerance of wild type Cupriavidus necator H16 to chemicals

To determine the tolerance level of wild type *C. necator* to target chemicals, the bacterium was cultivated in a 48-well microtiter plate containing 1 mL rich (LB) or defined media supplemented with different concentrations of ethanol, isopropanol, isobutanol, 3-methylbutanol, heptane and pentadecane. The cultivation was allowed to proceed up to 72 h in defined medium since the growth of bacterium peaks at 72 h when cultivated in such media, whilst growth in rich medium was measured after 24 h. The bacterium displayed similar response to each chemical regardless of the cultivation medium (Fig 5.2). However, growth under rich medium was more robust within the tolerance range of the bacterium to each chemical. The bacterium was able to tolerate up to 1% (v/v) ethanol in both growth media, above this concentration no considerable growth was observed (Fig 5.2A). The tolerance range of the bacterium to isopropanol, isobutanol and 3-methylbutanol were similar (Fig 5.2B–D). The bacterium's maximum tolerance level to isopropanol was 0.5% (v/v). At 0.5% (v/v) isobutanol or 3-methylbutanol, the bacterium grew slightly better in defined medium compared to rich medium, which showed no evidence of growth (Fig 5.2C and D). This is perhaps due to the longer incubation time in the defined medium. It appears the bacterium can tolerate 0.5% (v/v) hydrocarbons, specifically pentadecane (Fig 5.2E and F). In general, *C. necator* can tolerate minimal concentration of some of the chemicals tested. Thus, it is crucial to engineer the bacterium for improved tolerance to these chemicals.

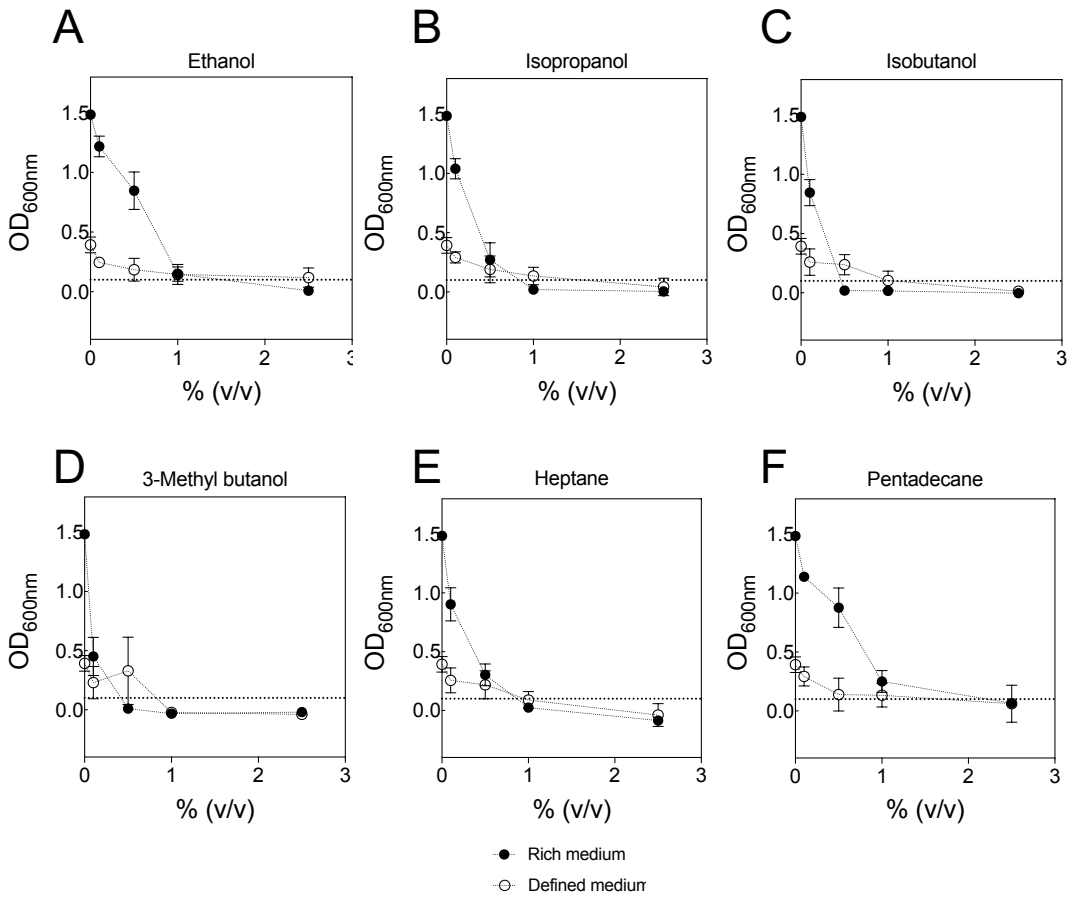


Fig. 5. 2 Growth of wild type *C. necator* H16 on chemicals.

The bacterium was tested against each chemical at 0, 0.1, 0.5, 1 and 2.5% (v/v). Growth was measured after 24 h cultivation in a rich medium (closed circles) or 72 in a defined medium (open circles). Horizontal broken line indicates 0.1 OD_{600nm} considered as growth.

5.3.2 Phenotype competitive growth assay starting at low concentrations of each chemical

Following construction and direct electroporation of *C. necator* with plasmids (restriction-ligation products) carrying mutated or unmutated *rpoD* genes, the bacterium was successfully transformed with very high efficiencies (Table 5.3). As expected, in all transformation events, pCAT201 (MCS) transformant had considerable higher transformation efficiency than *rpoD* transformants. In the first event, there was no significant difference (p = 0.38) in the transformation efficiency between mutated and unmutated *rpoD* for both genes, *H16_A1626* and *H16_A2725*, across the three mutation rates. Next, transformants from agar plates were suspended in 20 mL LB broth without any chemical to create a pooled liquid library and from this library, 1.5% (300 μL) was transferred into 20 mL LB-kanamycin further supplemented with low concentration of each chemical (Fig 5.3A). Depending on the level of growth, serial cultivation (using the same inoculum volume (300 μL)) was carried out starting from the least concentrated culture as described (Fig. 5.3A).

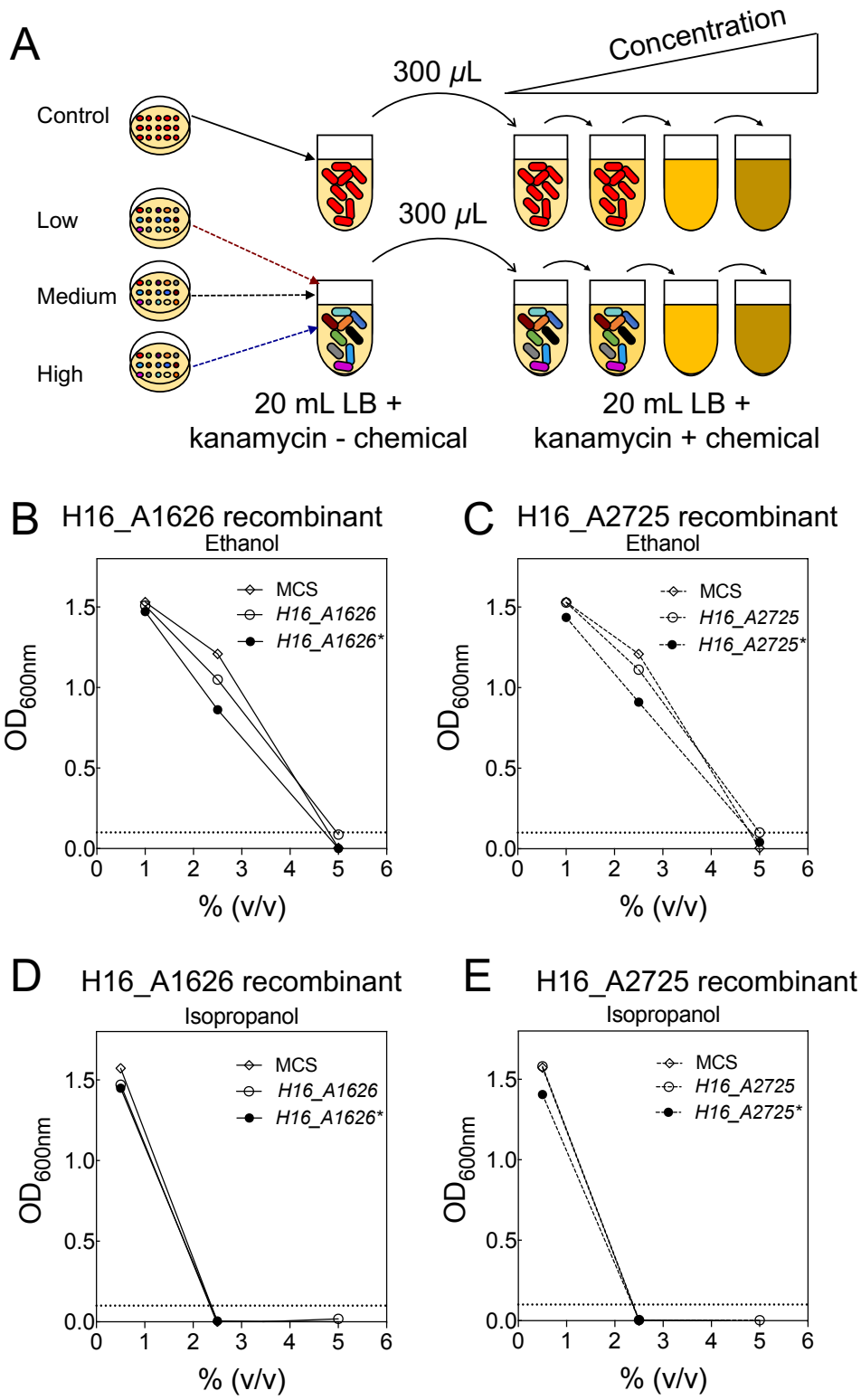


Fig. 5. 3 Phenotype screening of recombinant *C. necator* starting at low concentrations of each chemical.

A. Inoculation of 20 mL LB broth and serial cultivation in fresh broth with increasing concentration of each chemical. **B.** Tolerance of *H16_A1626* mutated and unmutated library to ethanol. **C.** Tolerance of *H16_A2725* mutated and unmutated library to ethanol. **D.** Tolerance of *H16_A1626* mutated and unmutated library to isopropanol. **E.** Tolerance of *H16_A2725* mutated and unmutated library to isopropanol. Plasmid pCAT201 with MCS (diamond symbols) served as the reference to distinguish between tolerance resulting from chromosomal mutation from that of plasmid borne mutation for mutated (closed circles) or unmutated (open circles) *rpoD*. Mutated *rpoD* is detonated with asterisk (*). Growth was measured at 48 h. Error bars are S.E.M., *n* = 3 replicates. Broken horizontal lines indicate 0.1 OD_{600nm} considered as growth.

Table 5. 3 Transformation efficiency of *C. necator* with plasmid habouring *rpoD.*

Plasmid ID	Size (bp)	T.E (1st event)	T.E (2nd event)
pCAT201	3198	$2.72 \times 10^6 \pm 3\%$	$2.72 \times 10^6 \pm 3\%$
pCATH16_A1626	5009	$5.46 \times 10^5 \pm 0.4\%$	$5.02 \times 10^5 \pm 4\%$
pCATH16_A1626*_low	5009	$6.05 \times 10^5 \pm 2\%$	$7.01 \times 10^5 \pm 9\%$
pCATH16_A1626*_medium	5009	$4.41 \times 10^5 \pm 1\%$	$5.30 \times 10^5 \pm 8\%$
pCATH16_A1626*_high	5009	$6.16 \times 10^5 \pm 1\%$	$5.06 \times 10^5 \pm 3\%$
pCATH16_A2725	5560	$5.00 \times 10^5 \pm 2\%$	$6.76 \times 10^4 \pm 35\%$
pCATH16_A2725*_low	5560	$5.73 \times 10^5 \pm 3\%$	$1.63 \times 10^5 \pm 32\%$
pCATH16_A2725*_medium	5560	$4.82 \times 10^5 \pm 1\%$	$6.87 \times 10^4 \pm 55\%$
pCATH16_A2725*_high	5560	$4.48 \times 10^5 \pm 2\%$	$5.16 \times 10^4 \pm 94\%$

Plasmids were assembled using a Golden gate method and delivered directly to *C. necator*. All transformation efficiencies (T.E.) were obtained as transformants/μg DNA. Percentage S.E.M., $n = 2$ replicates. Asterisk (*) denote plasmids with *rpoD* originating from random mutagenesis.

C. necator library resulting from mutated and unmutated *H16_A1626* and *H16_A2725* tolerated up to 2.5% (v/v) ethanol at 48 h cultivation (Fig. 5.3B and C), unlike during the previous small scale (1 mL) cultivation performed with the wild type. Nonetheless, MCS recombinant also survived at this concentration, 2.5% (v/v) ethanol. At 5% (v/v) ethanol all *C. necator* recombinants were unable to grow. Additionally, all recombinants displayed similar tolerance to isopropanol at the tested concentrations: 0.5, 2.5 and 5% (v/v). The recombinants were able to grow at 0.5% (v/v) isopropanol (Fig. 5.3D and E). None of the *rpoD* transformants were able to grow at 0.5% or higher concentration of isobutanol or 3-methylbutanol. Although the bacterium appeared to grow at higher (> 5%) heptane or pentadecane (OD_{600nm} > 1.55), it was observed that growth was between the LB interfaces, as the medium was not miscible with both hydrocarbons in 50 mL falcon tube. Due to a lack of evidence of growth at the lower concentration of isobutanol or 3-methylbutanol, and the immiscible nature of LB with heptane or pentadecane, these chemicals were excluded from further studies, with more attention focussed on ethanol and isopropanol.

5.3.3 Phenotype competitive growth assay starting at high concentration of ethanol and isopropanol

Based on the outcome of the competitive growth assay at small-scale cultivation, with initial cultivation at low concentration of each chemical, a further attempt was made to engineer *C. necator* to survive at higher concentration of these two chemicals, ethanol and isopropanol. In the first assay, it is possible that a mutant having the desired phenotypic trait was part of the transformants from which the liquid library was created, but that such a transformant was not carried through as inoculum due to the scale of the culture medium. To circumvent this, plasmids were reassembled using the same bioparts as with the first assembly. Like the first transformation event, *C. necator* was transformed with high efficiency (Table 5.3). However, unlike in the first transformation event, the transformation efficiency of *C. necator*

transformed with *H16_A2725** (mutated plasmid library) was considerably low ($p < 0.05$), and this decreased with increasing mutation rate. A liquid library was created as described previously for the small-scale competitive assay. However, to compensate for the low volume of inoculum used in the previous assay—and to increase the chance of mutant having the desired phenotype to outcompete those with undesired phenotype—the volume of culture medium was raised to 1 L, with the starting culture medium supplemented with higher concentration of chemical (5% and 3.5% ethanol and isopropanol, respectively) (Fig. 5.4A). These concentrations were previously established to be detrimental to both the wild type and *rpoD* recombinant *C. necator* (Fig. 5.2 and 5.3).

The results indicated that liquid library of *C. necator* transformants with mutated (*H16_A1626**) and unmutated (*H16_A1626*) *rpoD* were unable to grow at higher concentration of ethanol and isopropanol, 5 and 3.5% v/v, respectively (Fig. 5.4B). Intriguingly, liquid library of *C. necator* created from transformants with mutated (*H16_A2725**) *rpoD* grew at 5% v/v ethanol and 3.5% v/v isopropanol in the 1 L cultivation scale (Fig. 5.4C). The difference in the level of growth measured for ethanol was ~ 84% higher compared to that at small-scale assay at the same concentration. For both chemicals, OD_{600nm} was above the reference (0.1). Cognisant that the apparent tolerant *H16_A2725** mutant library with the desired phenotype might be too diluted in the 1 L culture, 1.5% (300 μL) inoculum from this culture was transferred into flask containing lower concentration of chemical (3% v/v ethanol or 2.5% isopropanol) to obtain higher culture density. These concentrations were already established to be detrimental to the wild type *C. necator*. Subsequently, instead of a serial cultivation, 1.5% (300 μL) inoculum was transferred from the flask containing the lower concentration of each chemical directly into flask containing higher concentrations of chemical, up 7% v/v ethanol (Fig. 5.4A). Growth improved at 5% (v/v) ethanol but decreased with increasing concentration (Fig. 5.4D). There was evidence of growth at 7% v/v ethanol, with growth higher than 0.1 OD_{600nm}. Further cultivation in isopropanol resulted in similar growth pattern between *H16_A2725** and MCS *C. necator* library (Fig. 5.4E). Nevertheless, growth from *C. necator H16_A2725** mutated library was clearly different from that of unmutated *H16_A2725* library. Due to a lack of clear difference in growth between *C. necator H16_A2725** and MCS culture in a medium containing up to 3.5% isopropanol, this chemical was not tested in subsequent experiments.

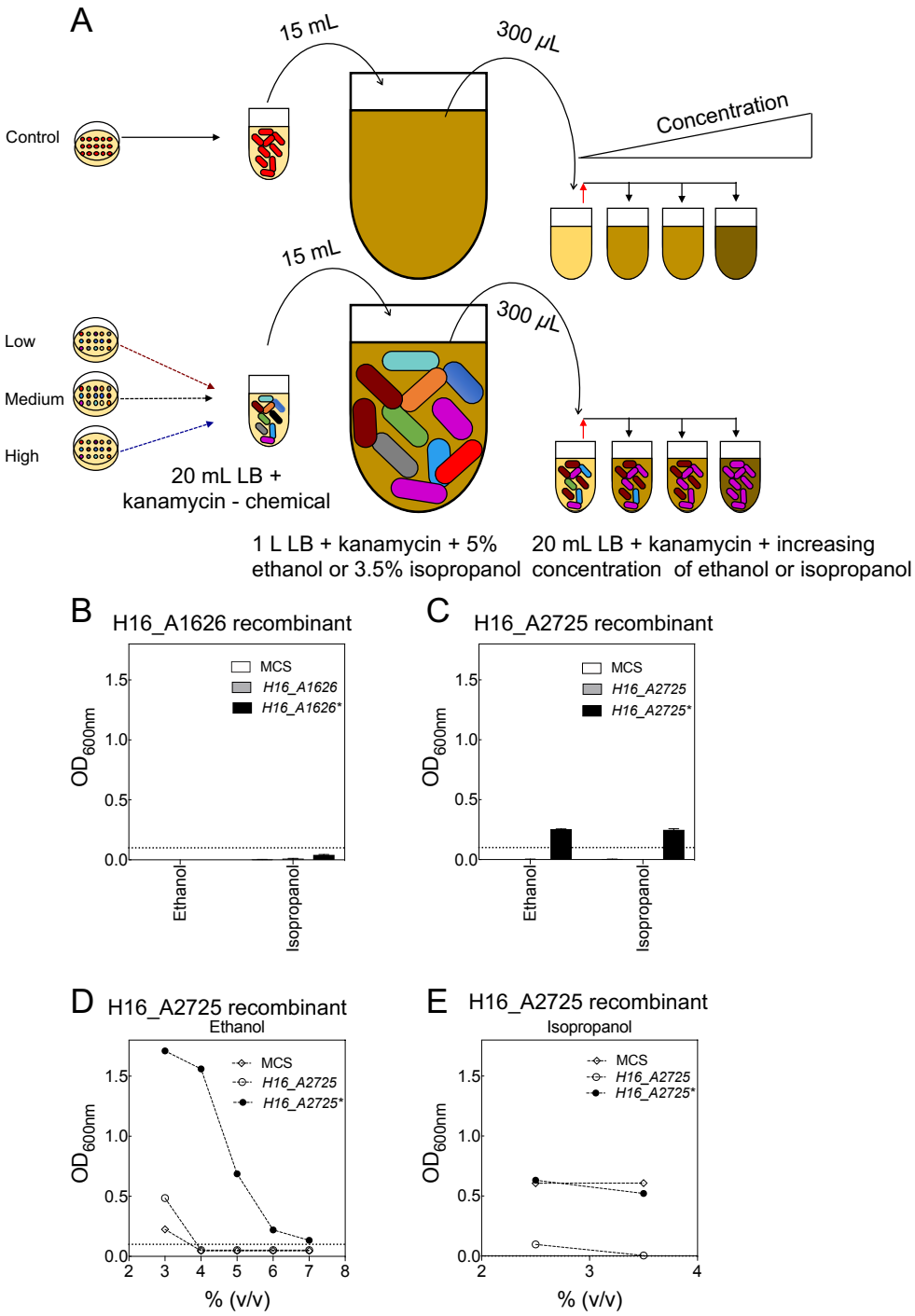


Fig. 5. 4 Phenotype screening starting at high concentration of ethanol and isopropanol.

A. Inoculation of 1 L LB broth supplemented with 5% (v/v) ethanol or 3.5% (v/v) isopropanol using 1.5% (15 mL) pooled liquid library as the inoculum. At 72 h incubation 1.5% (300 µL) culture was transferred from the 1 mL culture into flask containing 20 mL fresh LB medium supplemented with kanamycin and ethanol or isopropanol. **B.** Tolerance of *C. necator H16_A1626* mutated and unmutated library in medium containing 5% (v/v) ethanol and 3.5% (v/v) isopropanol at 1 L cultivation scale. **C.** Tolerance of *C. necator H16_A2725* mutated and unmutated library in medium containing 5% (v/v) ethanol and 3.5% v/v isopropanol at 1 L cultivation scale. **D.** Confirmation of *C. necator H16_A2725** mutant library tolerance to higher concentration of ethanol at 20 mL cultivation scale. **E.** Confirmation of *H16_A2725** mutant library tolerance to higher concentration of isopropanol. Plasmid pCAT201 with MCS (diamond symbols) served as the reference to distinguish between tolerance resulting from chromosomal mutation from that of plasmid borne mutation for unmutated (open circles) or mutated (closed circles) *rpoD*. Mutated *rpoD* is detonated with asterisk (*). Growth was measured at 48 h. Error bars are S.E.M., *n* = 3 replicates. Broken horizontal lines indicate 0.1 OD_{600nm} considered as growth.

5.3.4 Validating C. necator tolerance to high ethanol concentrations

Culture from tube containing 5% (v/v) ethanol and *C. necator H16_A2725** mutant library was spread on LB agar plate, supplemented with 200 *μ*g/mL kanamycin and incubated at 30°C for 48 h. This was to confirm that the observed tolerance of *C. necator* with mutated *H16_A2725** library—in the presence of high concentration of ethanol (> 5% v/v)—is the direct result of a mutant *H16_A2725**. The resulting colonies were cultivated in 48-well plate containing 1 mL LB supplemented with kanamycin and 5% (v/v) ethanol to screen and select transformants with best tolerance to the stress condition. Interestingly, individual colonies (eighteen) originating from *C. necator H16_A2725** library grew under the stress condition, 5% ethanol (Fig. 5.5A). Compared with the control colony with unmutated *H16_A2725*, the grow rates of the colonies originating from *H16_A2725* mutation were faster (Fig. 5.5A). Next, six cultures were selected and further cultivated at 20 mL culture scale, with medium containing 4–7% (v/v) ethanol. All cultures cultivated at this scale showed similar pattern of growth (Fig. 5.5B). Growth decreased with increasing concentration of ethanol.

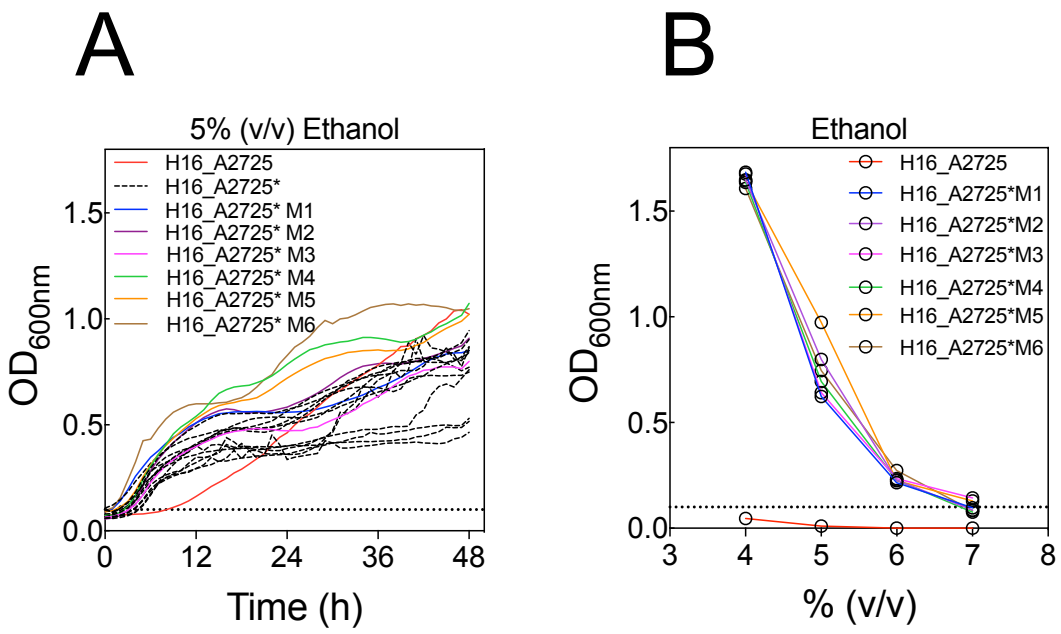


Fig. 5. 5 Validating *C. necator* library tolerance to ethanol.

A. High-throughput screening of eighteen colonies in 48-well microtiter plate containing 1 mL LB-kanamycin and 5% (v/v) ethanol. **B.** Scale-up cultivation of six selected cultures in 20 mL LB-kanamycin containing 5% (v/v) ethanol.

To further validate that the observed tolerance of *C. necator* to ethanol is plasmid mediated, plasmids purified from cultures derived from single colonies were delivered to wild type *C. necator*. Similarly, transformants were screened in 48-well microtiter plate containing 1 mL LB supplemented with kanamycin and 5% (v/v) ethanol. The growth of *C. necator* strains transformed with plasmids originating from mutated *rpoD* (*H16_A2725**) library were like that of strain with unmutated *rpoD* (*H16_A2725*) or MCS (Fig 5.6). In contrast to the previous observation, all the strains were unable to grow at high ethanol concentration (5% v/v ethanol) until after 12 h. Addionally, when each culture was transferred to 20 mL fresh LB-kanamycin medium (containing 5% (v/v) ethanol) in shake flask, none grew at 4–7% ethanol. It is more likely that the ability of the initial library (Fig. 5.5) to growth at high concentration of ethanol is not as a result of plasmid mediated *rpoD* mutation. This library likely evolves to tolerate high concentration of ethanol due to initial cultivation in high concentration of ethanol (Fig. 5.4).

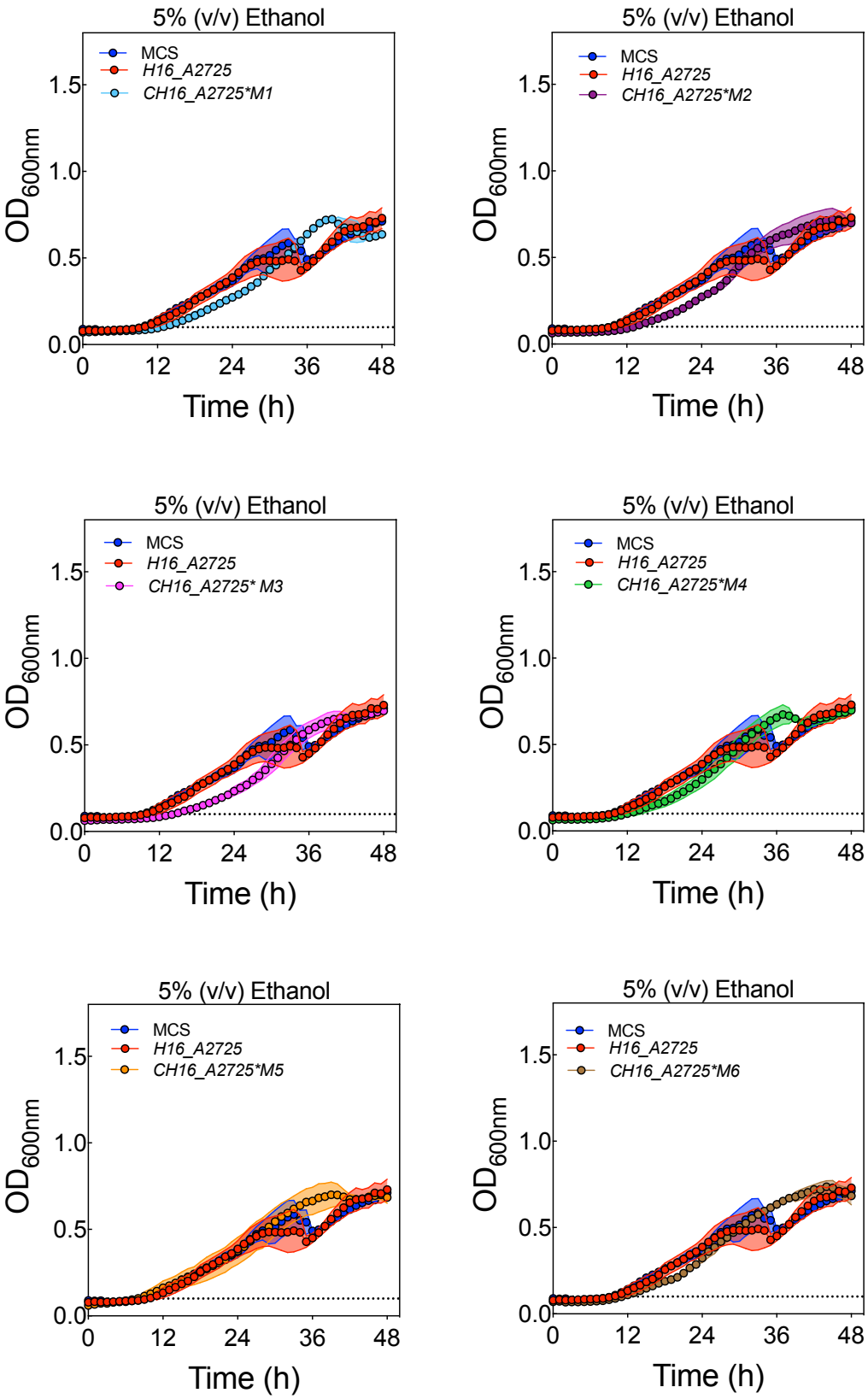


Fig. 5. 6 *C. necator* tolerance to ethanol is not plasmid mediated.

Plasmid derived from six culture tolerant to ethanol was electroporated into wild type *C. necator*. Three transformants from each was cultivated in 48-well plate containing 1 mL LB, 200 μg/mL kanamycin and 5% (v/v) ethanol. Plasmid pCAT201 with MCS (blue lines) served as the reference to distinguish between tolerance resulting from chromosomal mutation from that of plasmid borne mutation for unmutated (closed red circles) or mutated (closed circles with different colours) *rpoD*. Error bars are S.E.M., *n* = 3 biological replicates.

5.3.5 Challenges confirming mutation in rpoD sequence

It is established that *C. necator* tolerance to higher concentration of ethanol is not as a result of plasmid mediated (*H16_A2725**) *rpoD* mutation. However, attempts ($n > 3$) were made to sequence the *H16_A2725* rpoD* sequence originating from random mutagenesis. Plasmids from each culture (Fig. 5.5B) were purified and analysed on 0.8% agarose gel (Fig. 5.7A). Attempts to sequence the target, *H16_A2725**, were unsuccessful. Further, attempts ($n > 3$) to amplify the target by error-free PCR resulted in very low yields. For some of the plasmids, yields obtain following PCR were very low, whilst for others the target was not amplified (Fig. 5.7B). However, *KanR* which is part of the recombinant plasmid was amplified (Fig. 5.7C). Moreover, *H16_A2725* was also amplified from genomic DNA using the same primer set and Q5 DNA polymerase under identical thermocycling conditions (Fig. 5.7C). Therefore, the inability to successfully sequence *H16_A2725** from plasmid originating from random mutagenesis or amplify the sequence using error-free PCR is ascribed to the plasmid topology. It is possible that each colony (Fig. 5.5B) contains more than one plasmid. It is also possible that the *H16_A2725* rpoD* sequence in the recombinant plasmids is inaccessible to the primers resulting in low yield or lack of amplification for some of the plasmid templates. The plasmid DNA sequence peaks overlap and with some degree of noise thus cannot be relied on to identify mutation in the *H16_A2725** (Fig. 5.8). Nonetheless, the evidence of amplification when some of the plasmids were used as templates for a PCR (Fig. 5.7B) suggests differences in the mutated *H16_A2725** sequence in some of the plasmids. This apparent difference further indicates mutation in the *rpoD* (*H16_A2725**) sequence.

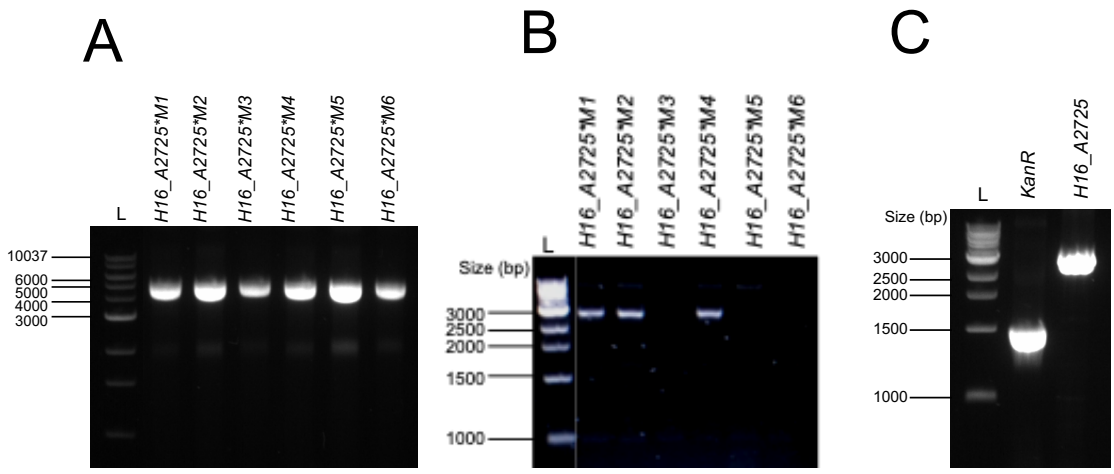


Fig. 5. 7 Plasmid and amplicon analyses for *rpoD*.

A. Gel electrophoresis of plasmids carrying *rpoD* originating from error-prone mutagenesis event. Plasmids were isolated from strains able to grow at 5% (v/v) ethanol. **B.** Error-free amplification of *rpoD* using plasmid recovered from ethanol tolerant cultures (Fig. 5.7A) as templates. **C.** Confirming the presence of kanamycin resistant cassette (*KanR*) in plasmid and amplification of *rpoD* from genomic DNA. L: 1kb ladder. Strain originating from mutated *rpoD* is designated with asterisk (*).

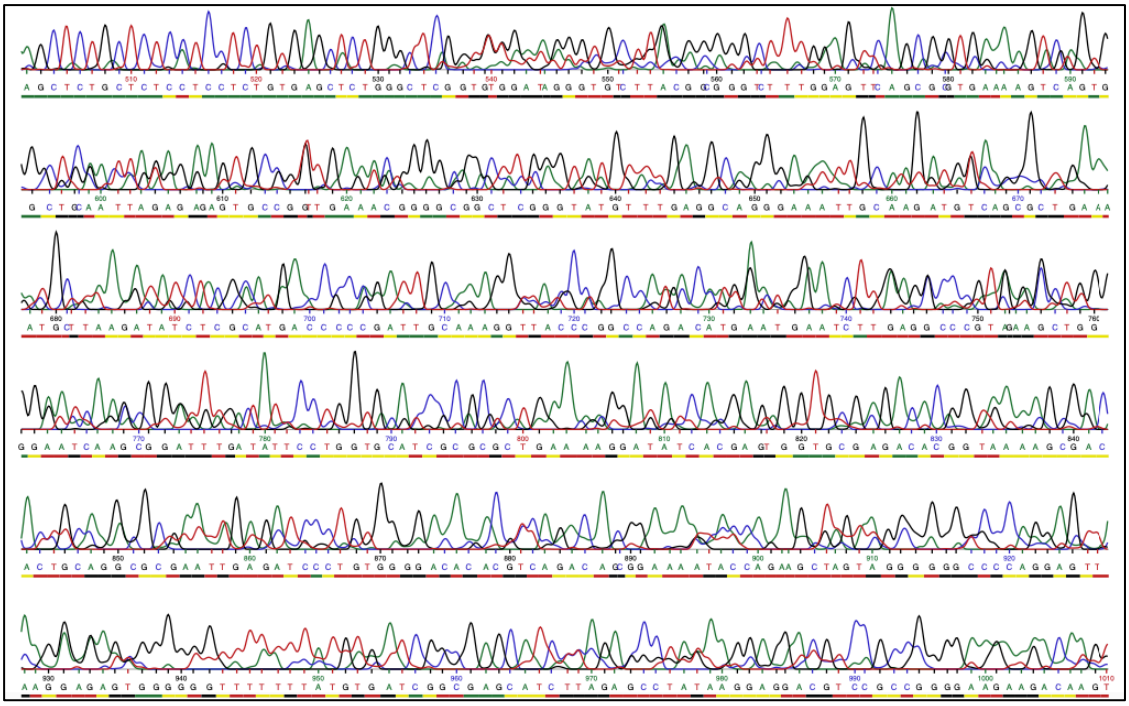


Fig. 5. 8 DNA chromatogram of *H16_A2725* rpoD*.

Number of attempts, *n* > 3.

5.4 Discussion

This study generated *Cupriavidus necator* strains able to grow at elevated concentration of ethanol. The bacterium has promising features for industrial biotechnological applications and has been used as a production platform for biosynthesis of platform chemicals [8,23]. Ethanol is one of the common and abundant biofuels [212]. Although bacteria are non-model microbes for ethanol production compared to *Saccharomyces cerevisiae*, the metabolic pathway of *C. necator* for ethanol production is simple, involving the conversion of acetyl-CoA to ethanol [139]. Given the ability of *C. necator* to stockpile carbon and interconvert it during growth, the bacterium has innate advantage to produce large quantity of ethanol with minimal engineering effort. The use of *C. necator* mutant unable to divert the stockpiled carbon to biopolymer accumulation favours the overproduction of precursor molecule for ethanol production. One of the factors limiting productivity in industrial biotechnology is toxicity of product to the production host [42]. Regardless of host metabolic network towards product biosynthesis, the use of host able to tolerate high concentration of product during biosynthesis offers an advantage to the production process. Such superior production host can be obtained by rational engineering or evolutionary approach.

At high concentration of alcohols, *C. necator* grew slightly better in a defined medium in comparison to growth in a rich medium. Although the bacterium was cultivated for a longer period in a defined medium, the slightly higher growth obtained is likely due to accumulation of polyhydroxybutyrate (PHB). Defined medium due to its composition favours PHB accumulation unlike complex medium, which contains undefined composition—with its components likely not limiting within shorter cultivation period to induce PHB accumulation. Cultivation of *C. necator* in a defined medium containing alcohols, specifically ethanol, enhances PHB accumulation [43]. It follows that if the bacterium is able to metabolise ethanol, the resulting pool of acetyl-CoA (precursor of PHB) together with reduced coenzymes, which in turn has the tendency of inhibiting tricarboxylic acid cycle (TCA) would support diverting acetyl-CoA to PHB biosynthetic pathway. Therefore, the use of rich medium is more suitable than defined medium for studies involving *C. necator* tolerance to alcohols.

Plasmid with MCS as the cargo (pCAT201) transformed *C. necator* with higher efficiency in comparison to variants with unmutated or mutated *rpoD*. This indicates that expressing *rpoD* might have elicited metabolic burden on the host. In the second transformation event (Table 5.3), the decrease in transformation efficiency obtained for pCATH16_A2725* with increasing mutation rate is ascribed to multiple mutation in the target. This observation is the first indication that mutation of the *rpoD* genes was perhaps achieved despite the inability to

successfully confirm such mutation by sequencing the target. In *C. necator*, it appears that a high mutation rate (9–16 mutation per kb) in *rpoD* (*H16_A2725**) might result in a desired phenotypic trait. However, such overall mutation (27–48 mutation rate in the *H16_A2725**) might also impact negatively on the replication function in the bacterium. In evolutionary studies, multiple mutations tend to inactivate proteins [213].

The slight increase in tolerance level obtained for ethanol (1–2.5%) and isopropanol (0.5%) during scale-up cultivation in shake flask (Fig. 5.3), compared to that obtained at microtiter scale (Fig. 5.2) with the wild type, is attributed to better growth condition and not as a result of strains harbouring *rpoD*—since similar growth pattern was obtain for pCAT201 transformants bearing MCS as a cargo. Although cultivation in microtiter enables high-throughput screening, the limited surface area for efficient mass transfer would likely reduce growth rate. The bacterium displayed high and consistent sensitivity to isobutanol and 3-methyl butanol. This is perhaps due to the less volatility of the compounds in comparison to those with lower molecular weight, ethanol and isopropanol.

This study demonstrates that cultivation of *C. necator* under high environmental stress condition led to strains able to withstand such stress. This evolutionary approach gained significant attention in the past two decades, and it is currently one of the approaches deployed by industrial laboratories to obtain superior host for production process [213,214]. The major challenge with such approach is the limitation in the number of mutants that can be screened [215]. As observed in this study the desired phenotypic trait, improved tolerance to ethanol, was obtained in the second iteration when the starting library and environmental stress were increased. The tolerant strains had faster growth rate when cultivated in 5% (v/v) ethanol in comparison to the control strain with unmutated *rpoD* (*H16_A2725*) (Fig. 5.5A). There are two possible explanations to this observation. First, the tolerance is as a result of ethanol metabolism. This implies that the tolerant strain imported the ethanol and metabolised it. The second possibility is that the tolerant strain used an efflux pump mechanism to prevent transportation of ethanol into the cell, which it turns improved growth in the presence of high concentration of ethanol. Efflux pump is one of the mechanisms employed by Gram-negative bacteria to survive in high concentration of solvents [197,198]. Regardless of the mechanism of tolerance, the result was improved tolerance of *C. necator* to ethanol, resulting in a short lag and fast exponential growth phase in LB containing 5% (v/v) ethanol. This study highlights the challenges of strain improvement for biotechnological applications. The constraint of transferring the desired phenotypic traits to another strain and identify and control the genotype leading to the desired phenotype still present challenges in this approach of strain engineering [199].

5.5 Conclusion

This study demonstrates that *C. necator* can be directed to evolve a new beneficial phenotypic trait. The evolved phenotypic trait is survival and faster growth rate in a rich medium supplemented with up to 7% (v/v) ethanol. This trait is crucial in the industrial application of *C. necator*, especially as a production platform for ethanol as a biofuel molecule. The use of a starting inoculum with high microbial population and cultivation starting at a high concentration of environmental stress favoured evolution of desired phenotype. High mutation rates result in low transformation efficiency. Although this study is unable to establish that mutation in plasmid-borne *rpoD* is linked to the desired phenotypic trait, the control strain, with ummutated *rpoD*, did not survive when cultivated in 5% (v/v) ethanol. Low concentration of higher alcohols (isobutanol and 3-methyl butanol) inhibited growth of *C. necator* indicating that future engineering effort to produce such compounds in the bacterium might be limited by the toxicity of these compounds to the host. In general, this study showed that *C. necator* can be engineered to tolerate high concentration of ethanol and can potentially be engineered to tolerate even higher concentrations. This approach can be deployed to generate *C. necator* tolerant to other industrial solvents thus making it a first-rate microbial chassis for industrial biotechnological applications.

Chapter 6.

Conclusions and Perspectives

6.1 General conclusion

The application of statistical design of experiments (DoE) in modelling *C. necator* growth in a defined medium demonstrates that such an approach can be deployed to gain useful information on the response of the bacterium towards different environmental conditions, with minimal experimental runs. The growth of the bacterium, despite its complex biochemical processes, can be predicted under different media composition. The growth model resulting from the DoE approach showed amino acids, fructose, $CaCl_2$, Na_2HPO_4 and trace elements contribute significantly to growth. Their individual effect on growth is concentration dependent. High concentration of amino acids, fructose and $CaCl_2$ favoured high growth output, whilst low concentration of Na_2HPO_4 and trace elements also favoured the same output. This indicates that defined medium supporting robust growth of *C. necator* will constitute less Na_2HPO_4 and trace elements. Based on the information obtained from the DoE approach, minimal growth media—containing only components that are essential or important to growth—can be formulated for different *C. necator* applications. Such media with only components contributing actively to growth would save cost, especially at large-scale cultivations, thus making *C. necator* growth medium economically competitive with growth media for other microbial chassis.

Although the output during the growth modelling was optical density (OD_{600nm}), the application of this approach can be extended to synthesis of high valued compounds in the bacterium. In this study, bioprocess parameters (pH, temperature, dO_2, agitation) were maintained at constant values guided by existing studies [14,15]. However, these parameters can change during fermentation, depending on the nature of product that is synthesised. Thus, varying these parameters during the modelling would likely impact on the model prediction and interpretation. Nonetheless, these parameters can be investigated during biosynthesis of a specific product to gain detailed understanding of bioprocess parameters on productivity.

The statistical DoE approach demonstrated in this study can be applied to quantitative proteomic analysis of *C. necator* under different media compositions. Proteomic analysis of *C. necator* have been carried out, however, with poorly characterised growth media [36,38,92].

With well-characterised media, for which the impact of every component on growth is known, differential expression of *C. necator* proteins across its three replicons (chromosome 1, chromosome 2 and pHG) will be better understood and interpreted. More so, this approach can be extended to quantitative fluxomic analysis of *C. necator* under autotrophic growth conditions with different media composition. The information obtained would guide rational pathway engineering in the bacterium, especially for the utilisation of cheap and renewable substrates. Similar quantitative flux analysis has been carried out for *C. necator* under heterotrophic and mixotrophic conditions [110]. Additionally, like in cyanobacteria [216] generating more information on *C. necator* fluxomics, proteomics and metabolomics would help build a kinetic mode that would potentially govern *C. necator* strain design.

C. necator can be transformed by heat-shock transformation. The major limitation of this new transformation protocol appears to be low plasmid yield. Perhaps, another plasmid backbone other than pBBR1 may be more suitable for optimising this transformation protocol in *C. necator*. With compatible plasmid backbone, DoE approach can be employed to gain insight into components that would have the most effect on *C. necator* heat-shock transformation efficiency. Electroporation has been optimised for *C. necator* using a one factor at a time approach [73,74]. This approach was unable to highlight interaction between the factors that were investigated. Beside the main effect of each factor considered during electroporation of *C. necator*, interaction between some of the factors (electroporation conditions or components of wash buffers) may result in further increase in transformation efficiency.

Tetracycline despite showing the highest growth inhibition is not the first choice as a selective pressure for selecting *C. necator* transformants. It follows that antibiotic inhibitory effect, exerted on non-recombinant strain, may not entirely be overcome in recombinant strain by coding resistance against some antibiotics. As observed for tetracycline, metabolic burden due to expression of antibiotic gene can impact negatively on plasmid propagation in a host. Antibiotic mode of action and susceptibility range are to be considered in choosing antibiotic cassette. The pBHR1 kanamycin cassette (*APH (3')-Ia*) remains the best antibiotic cassette for selecting *C. necator* transformants.

The mobilisation sequence of pBBR1 plasmids can be excluded from plasmid intended to be delivered to *C. necator* by electroporation. Similar adjustment was made for plasmid set intended for propagation in *Geobacillus* [217]. With the plasmid assembly method adopted in this study, Golden gate assembly, plasmids can rapidly be assembled and delivered directly to *C. nec*ator and the reporter genes expressed. This circumvents plasmid maintenance in *E. coli* and provides opportunity for phenotype prototyping in *C. necator*. A range of promoters

can drive gene expression in *C. necator*. Amongst the constitutive promoters tested in this study, P_{J23100}, with PET ribosomal binding site, constitutively induced the highest gene expression across reporter proteins (mRFP1 and eGFP). Although the plasmids built in this study were validated using constitutive promoters, these promoters can be substituted with inducible promoters to drive gene expression using this same pCAT plasmid backbone. The pCAT plasmid backbone is flexible thus allowing individual biopart to be sourced and used in constructing plasmids for other microbial chassis. More importantly, plasmid with more than one functional reporter, driven by different promoters, can be assembled efficiently using pCAT plasmid backbone. This indicates that multiplex genes can be assemble in a parallel fashion, efficiently delivered and expressed in *C. necator* using the compatible bioparts established in this study. Broadly, pCAT plasmid offers four efficiencies cloning approach: 1) it eliminates the reliance on conjugation as a means of delivering goi to *C. necator*; 2) it leads to high transformation efficiency in *C. necator* following electroporation; 3) it can be assembled and delivered directly to the destination host, without propagating in *E. coli*; 4) it can be propagated in a host over several generations without the addition of antibiotics.

C. necator has two *rpoD* genes, *H16_A1626* and *H16_A2725*, both differ in DNA and protein sequence. The latter has a non-canonical start codon, GTG (Appendix D). It is also not clear the function of each *rpoD* in the bacterium, whether both function under different environmental conditions. Further studies would shed light on the role of each *rpoD* and condition(s) that induces the regulatory function of one over another. Plasmid with mutant *H16_A2725** library evolved to survive in medium containing up to 7% (v/v) ethanol. Although the goal was to engineer *C. necator* to tolerate higher concentration of ethanol—as a result of plasmid induced tolerance at the global level—this study was unable to validate that the observe tolerance of the bacterium to ethanol is plasmid mediated. Given the inability to confirm mutation in the *H16_A2725**, it is not clear whether mutation of *H16_A2725** elicited other beneficial responses, pleiotropy, leading to *C. necator* survival in elevated concentration of ethanol in a culture medium. Due to the inability of the wild type *C. necator* to survival at 5% (v/v) ethanol when transformed with plasmid isolated from library able to growth at such ethanol concentration, it is more likely that the observed tolerance to ethanol evolved at the genome level. Plasmid curing of *C. necator* tolerant strains (Fig. 5.5), harbouring pCATH16_A2725* plasmid originating directly from the mutagenesis event, would enable direct comparison of these strains with that of the wild type. This would further establish if any mutation in the plasmid is linked to the survival of *C. necator* (H16_A2725*M1–M6) in medium containing 5% (v/v) ethanol. Further, transcriptional profiling of strains H16_A2725*M1–M6 might shed light on gene(s) that are differentially expressed

under both stress and unstressed conditions in comparisons to the wild type. This analysis can potentially reveal if the strains tolerant to ethanol are the same or different. Furthermore, genes that are upregulated or downregulated can be overexpressed or knockdown, respectively, to further establish that the identified gene are directly responsible for the observed phenotype, improved tolerance of *C. necator* to ethanol. Transcriptional analysis of ethanol tolerant *E. coli* and yeast strains was crucial in identifying key genes associated with improved tolerance to ethanol [201,202,210]. More importantly, the evolutionary approach demonstrated in this study can be applied to evolve *C. necator* with broad substrate utilisation and tolerance to other industrial solvents. The ability of *C. necator* H16 to utilise sugar is restricted to fructose and *N*-acetylglucosamine; however, prolonged cultivation in a medium with high glucose concentration resulted in a *C. necator* mutant able to utilise glucose as a carbon source [107]. Generating *C. necator* strain able to utilise cheap and more abundant substrates will contribute to making the bacterium an economically competitive industrial biocatalyst.

6.2 Significance of major findings based on hypotheses

- *C. necator* tends to have extended lag phase of growth (> 24 h) in a defined medium. Addition of few amino acids to *C. necator* defined growth medium improved growth rate and shortened lag phase of growth.

C. necator can synthesis all its essential cofactors *de novo*, including vitamins [9]. However, this study showed that some amino acids and trace elements are important supplements of *C. necator* growth medium. These supplements are not essential to growth but their addition increased growth rate and favoured robust growth. Specifically, histidine and $CuSO_4$ are key growth supplements that when added in the right proportion result in robust and reproducible growth. These two supplements are considered important. Copper is amongst the heavy metals of environmental concerns present in effluents [218–220]. It inhibits microbial growth at concentration found in wastewaters [219]. Therefore, *C. necator* is a potential good candidate, as a biosorbent, and can be deployed for treating wastewater consisting mostly of copper. The bacterium can be made to tolerate higher concentration of copper found in wastewater by adding histidine. This process is eco-friendlier compared to conventional wastewater treatment techniques, which are capitally intensive and result in sludge generation that require further disposal mechanisms [218,220].

How does copper stress affect intracellular histidine availability or biosynthesis, or how does the presence of sufficient histidine diminish the effect of copper stress? To address this question, two important scenarios from the bacterium's perspective need to be considered. Is the bacterium responding to copper toxicity or low availability of histidine, which is more detrimental to growth? Apparently, both scenarios are linked and are detrimental to growth. However, given that the bacterium can synthesis some of its amino acids sufficient to sustain growth [9], and copper established as an important trace metal for the bacterium's growth (Fig. 2.5), it is more likely that something else increases the importance of histidine under copper stress making histidine limitation detrimental to growth. It is noteworthy that it is excess unbound Cu, free copper ions (Cu^{+} or Cu^{2+}), that is detrimental to growth [221,222]. The toxicity of free Cu ions can range from catalysis of harmful redox reaction (when bound in weak sites), to disruption of enzyme functions (when bound in strong adventitious sites) [221]. Two P_{1B}-type ATPase subfamilies are known to export Cu ions. P_{1B-1}-type ATPase (CopAs) is a widely accepted Cu^{+} exporter, while P_{1B-3}-type ATPase (CopBs) selectivity on Cu ion varies among bacteria [223]. However, it is established recently that both enzymes are Cu^{+} exporters [223]. In yeast, it appears histidine does not directly chelate Cu^{2+} outside the cell environment, rather it interacts with Cu^{+} inside the cell to enhance its uptake thus reducing its availability in the cell [224].

Histidine is a strong metal coordinating tridentate ligand with three potential metal-binding sites: the carboxylate oxygen group, the imidazole imido group and the amino nitrogen group [225]. The binding to these three sites depends on the Cu ions and the complexes formed [225]. Bidentate coordination is favoured with *L*-histidine-Cu^{2+}; however, tridentate coordination is favoured under mixed *L*-amino acids-Cu^{2+} complexes. It appears copper (Cu^{+} or Cu^{2+}) stress does not impair histidine biosynthesis, rather it increases the demand of histidine in the cell. This is supported by an increased concentration of histidine and other amino acids in the xylem sap of *Brassica carinata* under condition of excess copper [226]. It is unclear which of the Cu ions toxicity is diminished by adding histidine. Answering this question will require robust metallomic studies of *C. necator*, which will help shed light on the selectivity of CopA and CopB on Cu ions in the bacterium. Further, it will help establish whether the toxicity is inside the cell (i.e. Cu^{+} toxicity) or outside the cell (i.e. Cu^{2+} toxicity). In summary, a simple model of copper and histidine interaction follows that: high concentration of free Cu ions signals or activates P_{1B}-type ATPase activity, which in turn depletes histidine availability making it a limiting amino acid for Cu ion exportation. It appears the binding is between copper and copper chaperones [227]. The latter play a critical role in delivering copper to P_{1B}-type ATPase by interacting with the N- (histidine rich) terminal of Cu-binding domain of P_{1B}-type ATPase [228]. It is still unclear the exact role histidine plays, and with which form of Cu ions, and where

in the cell this role is performed. What is clear is that histidine acts like an enhancer or promoter or coordinator of P_{1B}-type ATPase Cu ions exportation to maintain copper homeostasis. Whether histidine interacts with CopA or CopB, or both, and/or even an exporter yet to be identified in *C. necator* is yet to be clearly established.

- *C. necator* is susceptible to transformation with limited number of existing plasmids. Moreover, the transformation efficiencies obtained with exiting BHR plasmids are low. Incompatibility of bioparts is responsible for the low transformation efficiency observed for *C. necator* with existing plasmids.

pBHR1 has incomplete replication sequence, which impacted negatively on plasmid propagation in *C. necator*. Whilst pBBR1MCS-2 has complete replication sequence, its *KanR* (*APH (3')-IIa*) does not code enough resistance following electroporation into *C. necator*. pBBR1 *oriV* or *Rep* can independently direct autonomous replication of plasmids in *C. necator*. Nonetheless, replication with both parts results in higher transformation efficiency, stability, and expression of reporter proteins. This raises the question whether either of the part is what it has been ascribed. They might be a single replication module. pBBR1 does not belong to any of the known incompatibility groups [49,51]. Further studies on pBBR1 replication sequence would provide useful information especially on the incompatibility group, which might shed light on the replication function.

While it was possible to incorporate high copy *ori* to improve plasmid yield, such incorporation affected reporter protein expression. There seem to be a trade-off between expression of reporter proteins and high plasmid yield. Bacteria due to limited resources allocated for gene expression often trade-off between maximum growth, phenotype and gene expression [229]. Thus, the use of high copy *ori* may not bode well for gene expression in *C. necator* owing to metabolic burden elicited by expressing high copy number *ori* in *C. necator*. A balance therefore must be established between high copy number in the bacterium and heterologous gene expression. This understanding is crucial in designing complex circuits and pathways for heterologous gene expression in *C. necator*.

- Re-engineering of existing plasmids improved transformation efficiency in *C. necator*.

Significant improvement in transformation efficiency, segregational stability and expression of reporter proteins was observed when a broad host range (BHR) plasmid was re-engineered

for application in *C. necator* using compatible bioparts. The re-engineering results in BHR plasmid (pCAT) smaller in size than the existing BHR plasmids. The resulting pCAT plasmids can stably be propagated for more than six generations without the addition of antibiotic in a culture medium and still retain their ability to express reporter proteins. The ability to directly electroporate restriction-ligation reaction to *C. necator* and express more than one reporter protein carried on a single plasmid provides an opportunity for efficient metabolic engineering applications in the bacterium. Components of a metabolic pathway can be investigated and assembled in a combinatorial fashion to build a statistical model that covers an entire pathway productivity range. This multivariate-modular pathway engineering approach has been successfully demonstrated in *E. coli* for the biosynthesis of taxadiene [230], an important intermediate of taxol (an anticancer drug), violacein [97], fatty acids [231], (2S)-pinocembrin [232], and in *S. cerevisiae* for terminal alkene production [233]. Furthermore, with an improved transposon mutagenesis, *C. necator* viable minimal genome—with essential genes for a biochemical process of interest—can be designed, synthesised and housed in another receptive host. This approach has been deployed in building a viable *Mycoplasma mycoides* minimal genome, which contains approximately half (49.21% kb) the genome size of the original genome [87].

6.3 Future directions

Developing more experimental resources for C. necator

This study provides an update on the existing experimental resources and their applications in *C. necator*. It also highlighted some limitations that need to be addressed to make the bacterium more appealing for academic and industrial laboratory studies. Specifically, progress have been made on promoter engineering in the bacterium for controlled gene expression [57,62,63]. Other important bioparts that require extensive characterisation are translational (riboswitch and ribozyme switch) and post-transcriptional regulators. This would enable rational design and construction of a complex circuit to implement, possibly, a new function in *C. necator*. Genetic circuits are currently available for model microbial chassis from the early 2000's [234–236].

Propagation of new functions in *C. necator* rely mostly on disrupting an existing function(s) in the bacterium [11,12,77,79,108]. This is widely achieved by gene integration or more widely homologous recombination, which abolishes an existing gene. The process of recombination in *C. necator* is achieved using a suicide vectors, which are constructed using a narrow host range *ori* and other essential bioparts needed to achieve gene recombination or integration [76]. The first challenge of the protocol is selecting *C. necator* transformants over *E. coli*

following conjugation. Given that narrow host range *ori* cannot direct replication in *C. necator*, suicide vectors, with or without *ori*, can be assembled for the bacterium and delivered directly via electroporation. This would improve the efficiency of the protocol. Secondly, counter-selection of double crossover recombinants relies on their susceptibility to *Bacillus subtilis* levansucrase, *SacB*, in a rich medium containing 10% sucrose. This counter-selection tends to be ineffective at 10% sucrose leading to high rate of false positive recombinants [76]. Increasing the concentration of sucrose up to 25% reduced the chances of false positive recombinant [237]. As with *Clostridium* [238], identifying alternative counter-selectable marker would reduce the dependence on *SacB* as a tool for selecting *C. necator* recombinants. Additionally, guided gene recombination in *C. necator* can be achieved using CRISPR system [56]. This would help expedite the generation of knockout strains of *C. necator* by relying on fewer rounds of screening and selection.

Production of biopolymer with enhanced performance

Short chain-length polyhydroxyalkanoates (PHAs), like polyhydroxybutyrate (PHB), tend to have low glass transition temperature ($\mathbf{T_g}$), which determines other material properties of a biopolymer [239]. Incorporation of aromatic units into PHA from *C. necator* will increase the material features of polymer produced from the bacterium. With the tools established, future bioengineering of *C. necator* to produce biopolymer with the desired mechanical, rheological and thermal properties are in prospect. The group of native biopolymers (PHA) from *C. necator* has less annual production (30,000 tons) rate than other commercial biopolymers [239]. *C. necator* remains an ideal microbial candidate for the production of biopolymers owing to its ability to accumulate > 80% of its dry cell weight. Therefore, pathways for the production of the monomers of some conventional plastics (polylactic acid and polyethylene), which have wider applications can be expressed in C. *necator*.

Microbial ethylene biosynthetic pathway can potentially be expressed in *C. necator*. Ethylene biosynthesis in bacteria requires an ethylene forming enzyme (EFE)—from *Pseudomonas syringae* pv. *phaseolicola* PK2—to catalyse the formation of ethylene in a single step reaction [240,241]. This EFE has been successfully expressed in *E. coli* [242], Cyanobacteria [243] and yeast [244] for the production of ethylene. The precursor for microbial ethylene biosynthesis is an acetyl-CoA [240]. The use of *C. necator* strain unable to divert its acetyl-CoA to PHB will contribute to increased pool of α-ketoglutarate, the substrate for ethylene biosynthesis, in the TCA cycle. Arginine and oxygen are also key players during ethylene biosynthesis [240]. More importantly, ethylene can be produced in *C. necator* using CO_2 as the carbon source, unlike in *E. coli* and yeast; this will significantly reduce production cost and carbon footprint.

Pyridinedicarboxylic acid (PDCA), analogous of terephthalic acid—a building block of polybutyrate adipate terephthalate (PBAT)—has been produced in *Rhodococcus jostii* RHA1 from vanillic acid [245]. The production of PDCA relies on the presence of ammonia in a culture medium for cyclisation of ring cleavage product [245]. The genus *Cupriavidus* is a known degrader of aromatic compounds [9,84]. Thus, similar pathways for the degradation of some aromatic compounds in *C. necator* can be redirected, with minimal metabolic engineering, to produce aromatic building blocks of polymers with high performance properties. With the understanding of the impact of arginine and NH_4Cl on *C. necator* growth, cultivation media for the production of ethylene or pyridinedicarboxylic acid (PDCA) can readily be optimised.

Drug precursor molecules

Acetyl-CoA is one of the key precursors for natural occurring drugs [246]. With the innate ability of *C. necator* to divert its pool of acetyl-CoA to PHB under growth limiting condition, the bacterium can be engineered to redirect its acetyl-CoA into cytosol. Similar approach has been demonstrated in yeast with ethanol as the sole carbon source [246]. Moreover, instead of ethanol as the carbon source, CO_2 can be fed to *C. necator* as the sole carbon source to produce acetyl-CoA to be used as a drug precursor molecule thus reducing the production cost attributed to substrate.

The tools established in this study would be crucial in advancing future bioengineering applications in *C. necator* therefore making the bacterium a more propitious microbial chassis to be deployed to meet industrial and environmental needs.

References

1. Adams, B. L. The next generation of synthetic biology chassis: Moving synthetic biology from the laboratory to the field. *ACS Synth. Biol.* **5**, 1328–1330 (2016).
2. Calero, P. & Nikel, P. I. Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms. *Microb. Biotechnol.* **12**, 98–124 (2019).
3. Horwitz, A. A. *et al.* Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas. *Cell Syst.* **1**, 88–96 (2015).
4. Jakočiunas, T., Jensen, M. K. & Keasling, J. D. CRISPR/Cas9 advances engineering of microbial cell factories. *Metab. Eng.* **34**, 44–59 (2016).
5. Yaguchi, A., Spagnuolo, M. & Blenner, M. Engineering yeast for utilization of alternative feedstocks. *Curr. Opin. Biotechnol.* **53**, 122–129 (2018).
6. Nybo, S. E., Khan, N. E., Woolston, B. M. & Curtis, W. R. Metabolic engineering in chemolithoautotrophic hosts for the production of fuels and chemicals. *Metab. Eng.* **30**, 105–120 (2015).
7. Schmidt, M. *et al.* Poly(3-hydroxybutyrate-co-3-hydroxyvalerate) production in a system with external cell recycle and limited nitrogen feeding during the production phase. *Biochem. Eng. J.* **112**, 130–135 (2016).
8. Raberg, M., Volodina, E., Lin, K. & Steinbüchel, A. Ralstonia eutropha H16 in progress: Applications beside PHAs and establishment as production platform by advanced genetic tools. *Crit. Rev. Biotechnol.* **38**, 494–510 (2018).
9. Pohlmann, A. *et al.* Genome sequence of the bioplastic-producing 'Knallgas' bacterium Ralstonia eutropha H16. *Nat. Biotechnol.* **24**, 1257–62 (2006).
10. Cramm, R. Genomic view of energy metabolism in Ralstonia eutropha H16. *J. Mol. Microbiol. Biotechnol.* **16**, 38–52 (2008).
11. Li, H. *et al.* Integrated electromicrobial conversion of CO2 to higher alcohols. *Science (80-.).* **335**, 1596 (2012).
12. Lu, J., Brigham, C. J., Gai, C. S. & Sinskey, A. J. Studies on the production of branched-chain alcohols in engineered Ralstonia eutropha. *Appl. Microbiol. Biotechnol.* **96**, 283–297 (2012).
13. Oda, T. *et al.* Hydrogen-driven asymmetric reduction of hydroxyacetone to (R)-1,2-propanediol by Ralstonia eutropha transformant expressing alcohol dehydrogenase from Kluyveromyces lactis. *Microb. Cell Fact.* **12**, 2 (2013).
14. Grousseau, E., Lu, J., Gorret, N., Guillouet, S. E. & Sinskey, A. J. Isopropanol production with engineered Cupriavidus necator as bioproduction platform. *Appl. Microbiol. Biotechnol.* **98**, 4277–4290 (2014).

15. Marc, J. *et al.* Over expression of GroESL in Cupriavidus necator for heterotrophic and autotrophic isopropanol production. *Metab. Eng.* **42**, 74–84 (2017).
16. Müller, J. *et al.* Engineering of Ralstonia eutropha H16 for autotrophic and heterotrophic production of methyl ketones. *Appl. Environ. Microbiol.* **79**, 4433–4439 (2013).
17. Chen, J. S. *et al.* Production of fatty acids in *Ralstonia eutropha* H16 by engineering *β* -oxidation and carbon storage. *PeerJ* **3**, e1468 (2015).
18. Przybylski, D. *et al.* Exploiting mixtures of H2,CO2,and O2 for improved production of methacrylate precursor 2-hydroxyisobutyric acid by engineered Cupriavidus necator strains. *Appl. Microbiol. Biotechnol.* **99**, 2131–2145 (2015).
19. Crépin, L., Lombard, E. & Guillouet, S. E. Metabolic engineering of Cupriavidus necator for heterotrophic and autotrophic alka (e) ne production. *Metab. Eng.* **37**, 92–101 (2016).
20. Aneja, K. K., Ashby, R. D. & Solaiman, D. K. Y. Altered composition of Ralstonia eutropha poly(hydroxyalkanoate) through expression of PHA synthase from Allochromatium vinosum ATCC 35206. *Biotechnol. Lett.* **31**, 1601–1612 (2009).
21. Zhang, Y., Liu, G., Weng, W., Ding, J. & Liu, S. Engineering of Ralstonia eutropha for the production of poly (3-hydroxybutyrate- co -3-hydroxyvalerate) from glucose. *J. Biotechnol.* **195**, 82–88 (2015).
22. Barnard, G. C., Henderson, G. E., Srinivasan, S. & Gerngross, T. U. High level recombinant protein expression in Ralstonia eutropha using T7 RNA polymerase based amplification. *Protein Expr. Purif.* **38**, 264–271 (2004).
23. Chakravarty, J. & Brigham, C. J. Solvent production by engineered Ralstonia eutropha: channeling carbon to biofuel. *Appl. Microbiol. Biotechnol.* **102**, 5021–5031 (2018).
24. Heinrich, D., Raberg, M. & Steinbüchel, A. Studies on the aerobic utilization of synthesis gas (syngas) by wild type and recombinant strains of Ralstonia eutropha H16. *Microb. Biotechnol.* **11**, 647–656 (2018).
25. Wilde, E. Untersuchungen über Wachstum und Speicherstoffsynthese von Hydrogenomonas. *Arch. Mikrobiol.* **43**, 109–137 (1962).
26. Repaske, R. Nutritional requirements for Hydrogenomonas eutropha. *J. Bacteriol.* **83**, 418–22 (1962).
27. Davis, D. H., Doudoroff, M., Stanier, R. Y. & Mandel, M. Proposal to reject the genus Hydrogenomonas: taxonomic implications. *Int. J.* **19**, 375–390 (1969).
28. Yabuuchi, E., Kosako, Y., Yano, I., Hotta, H. & Nishiuchi, Y. Transfer of Two Burkholderia and An Alcaligenes Species to Ralstonia Gen. Nov. Proposal of Ralstonia pickettii (Ralston, Palleroni and Doudoroff 1973) Comb. Nov., Ralstonia

solanacearum (Smith 1869) Comb. Nov. and Ralstonia eutropha (Davis 1969) Comb. Nov. *Microbiol. Immunol.* **39**, 897–904 (1995).

29. Vaneechoutte, M., Kämpfer, P., De Baere, T., Falsen, E. & Verschraegen, G. Wautersia gen. nov., a novel genus accomodating the phylogenetic lineage including Ralstonia eutropha and related species, and proposal of Ralstonia [Pseudomonas] syzygii (Roberts et al. 1990) comb. nov. *Int. J. Syst. Evol. Microbiol.* **54**, 317–327 (2004).
30. Vandamme, P. & Coenye, T. Taxonomy of the genus Cupriavidus: A tale of lost and found. *Int. J. Syst. Evol. Microbiol.* **54**, 2285–2289 (2004).
31. Makkar, N. S. & Casida, L. E. Cupriavidus necator gen. nov., sp. nov.; A nonobligate bacterial predator of bacteria in soil. *Int. J. Syst. Bacteriol.* **37**, 323–326 (1987).
32. Schwartz, E., Gerischer, U. & Friedrich, B. Transcriptional regulation of Alcaligenes eutrophus hydrogenase genes. *J. Bacteriol.* **180**, 3197–3204 (1998).
33. Schwartz, E. *et al.* Complete nucleotide sequence of pHG1: A Ralstonia eutropha H16 megaplasmid encoding key enzymes of H2-based lithoautotrophy and anaerobiosis. *J. Mol. Biol.* **332**, 369–383 (2003).
34. Schwartz, E. & Friedrich, B. A physical map of the megaplasmid pHG1, one of three genomic replicons in Ralstonia eutropha H16. *FEMS Microbiol. Lett.* **201**, 213–219 (2001).
35. Ledger, T. *et al.* The complete multipartite genome sequence of Cupriavidus necator JMP134 , a versatile pollutant degrader. *PLoS One* **5**, e9729 (2010).
36. Schwartz, E. *et al.* A proteomic view of the facultatively chemolithoautotrophic lifestyle of Ralstonia eutropha H16. *Proteomics* **9**, 5132–5142 (2009).
37. Hogrefe, C., Romermann, D. & Friedrich, B. Alcaligenes eutrophus hydrogenase genes (Hox). *J. Bacteriol.* **158**, 43–48 (1984).
38. Kohlmann, Y. *et al.* Analyses of soluble and membrane proteomes of Ralstonia eutropha H16 reveal major changes in the protein complement in adaptation to lithoautotrophy. *J. Proteome Res.* **10**, 2767–2776 (2011).
39. Bartha, R. & Ordal, E. J. Nickel-dependent chemolithotrophic growth of two hydrogenomonas strains. *J. Bacteriol.* **89**, 1015–1019 (1965).
40. Friedrich, C. G., Friedrich, B. & Bowien, B. Formation of enzymes of autotrophic metabolism during heterotrophic growth of Alcaligenes eutrophus. *J. Gen. Microbiol.* **122**, 69–78 (1981).
41. Davis, D. H., Doudoroff, M., Stanier, R. Y. & Mandel, M. 9 47 20. *Int. J.* **19**, 375–390 (1969).
42. Marc, J., Grousseau, E., Lombard, E., Sinskey, A. J. & Gorret, N. Over expression of GroESL in Cupriavidus necator for heterotrophic and autotrophic isopropanol

production. *Metab. Eng.* **42**, 74–84 (2017).

43. Obruca, S., Marova, I. & Stankova, M. Effect of ethanol and hydrogen peroxide on poly (3-hydroxybutyrate) biosynthetic pathway in Cupriavidus necator H16. *World J. Microbiol. Biotechnol.* **26**, 1261–1267 (2010).
44. Lu, J., Brigham, C. J., Rha, C. & Sinskey, A. J. Characterization of an extracellular lipase and its chaperone from Ralstonia eutropha H16. *Appl. Microbiol. Biotechnol.* **97**, 2443–2454 (2013).
45. Li, H. *et al.* Integrated Electromicrobial Conversion of CO 2 to Higher Alcohols. *Science (80-.).* **335**, 1596 (2012).
46. Gruber, S., Schwab, H. & Koefinger, P. Versatile plasmid-based expression systems for Gram-negative bacteria-General essentials exemplified with the bacterium Ralstonia eutropha H16. *N. Biotechnol.* **32**, 552–558 (2015).
47. Jain, A. & Srivastava, P. Broad host range plasmids. *FEMS Microbiol. Lett.* **348**, 87–96 (2013).
48. Kovach, M. E. *et al.* Four new derivatives of the broad host range cloning vector PBBR1MCS, carrying different antibiotic resistance cassettes. *Gene* **166**, 175–176 (1995).
49. Szpirer, C. Y., Faelen, M. & Couturier, M. Mobilization function of the pBHR1 plasmid, a derivative of the broad-host-range plasmid pBBR1. *J. Bacteriol.* **183**, 2101–2110 (2001).
50. Bi, C. *et al.* Development of a broad-host synthetic biology toolbox for Ralstonia eutropha and its application to engineering hydrocarbon biofuel production. *Microb. Cell Fact.* **12**, 107 (2013).
51. Bi, C. *et al.* Development of a broad-host synthetic biology toolbox for ralstonia eutropha and its application to engineering hydrocarbon biofuel production. *Microb. Cell Fact.* **12**, 1–10 (2013).
52. Gruber, S., Hagen, J., Schwab, H. & Koefinger, P. Reprint of 'Versatile and stable vectors for efficient gene expression in Ralstonia eutropha H16'. *J. Biotechnol.* **192**, 410–418 (2014).
53. Solaiman, D. K. Y., Swingle, B. M. & Ashby, R. D. A new shuttle vector for gene expression in biopolymer-producing Ralstonia eutropha. *J. Microbiol. Methods* **82**, 120–123 (2010).
54. Hanko, E. K. R., Minton, N. P. & Malys, N. Characterisation of a 3-hydroxypropionic acid-inducible system from Pseudomonas putida for orthogonal gene expression control in Escherichia coli and Cupriavidus necator. *Sci. Rep.* **7**, 1–13 (2017).
55. Black, W. B., Zhang, L., Kamoku, C., Liao, J. C. & Li, H. Rearrangement of coenzyme A-acylated carbon chain enables synthesis of isobutanol via a novel pathway in

Ralstonia eutropha. *ACS Synth. Biol.* **7**, 794–800 (2018).

56. Xiong, B. *et al.* Genome editing of Ralstonia eutropha using an electroporation-based CRISPR-Cas9 technique. *Biotechnol. Biofuels* **11**, 1–9 (2018).
57. Johnson, A. O., Gonzalez-Villanueva, M., Tee, K. L. & Wong, T. S. An engineered constitutive promoter set with broad activity range for Cupriavidus necator H16. *ACS Synth. Biol.* **7**, 1918–1928 (2018).
58. Fukui, T., Ohsawa, K., Mifune, J., Orita, I. & Nakamura, S. Evaluation of promoters for gene expression in polyhydroxyalkanoate- producing Cupriavidus necator H16. *Appl. Microbiol. Biotechnol.* **89**, 1527–1536 (2011).
59. Gruber, S., Hagen, J., Schwab, H. & Koefinger, P. Reprint of 'Versatile and stable vectors for efficient gene expression in Ralstonia eutropha H16'. *J. Biotechnol.* **192**, 410–418 (2014).
60. Moore, S. J. *et al.* EcoFlex: A multifunctional MoClo kit for E. coli synthetic biology. *ACS Synth. Biol.* **5**, 1059–1069 (2016).
61. Gruber, S. *et al.* Design of inducible expression vectors for improved protein production in Ralstonia eutropha H16 derived host strains. *Journal of Biotechnology* **235**, 92–99 (2016).
62. Alagesan, S. *et al.* Functional genetic elements for controlling gene expression in Cupriavidus necator H16. *Appl. Environ. Microbiol.* **84**, 1–17 (2018).
63. Aboulnaga, E. A., Zou, H., Selmer, T. & Xian, M. Development of a plasmid-based, tunable, tolC-derived expression system for application in Cupriavidus necator H16. *J. Biotechnol.* **274**, 15–27 (2018).
64. Jugder, B.-E., Welch, J., Braidy, N. & Marquis, C. P. Construction and use of a *Cupriavidus necator* H16 soluble hydrogenase promoter (P_{SH}) fusion to *gfp* (green fluorescent protein). *PeerJ* **4**, e2269 (2016).
65. Arikawa, H. & Matsumoto, K. Evaluation of gene expression cassettes and production of poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) with a fine modulated monomer composition by using it in Cupriavidus necator. *Microb. Cell Fact.* **15**, 1–11 (2016).
66. Gruber, S. *et al.* Design of inducible expression vectors for improved protein production in Ralstonia eutropha H16 derived host strains. *J. Biotechnol.* **235**, 92–99 (2016).
67. Li, H. & Liao, J. C. A synthetic anhydrotetracycline-controllable gene expression system in ralstonia eutropha H16. *ACS Synth. Biol.* **4**, 101–106 (2015).
68. Lütte, S. *et al.* Autotrophic production of stable-isotope-labeled arginine in Ralstonia eutropha strain H16. *Appl. Environ. Microbiol.* **78**, 7884–7890 (2012).
69. Voss, I. & Steinbüchel, A. Application of a KDPG-aldolase gene-dependent addiction system for enhanced production of cyanophycin in Ralstonia eutropha strain H16.

Metab. Eng. **8**, 66–78 (2006).

70. Sobecky, P. A., Easter, C. L., Bear, P. D. & Helinski, D. R. Characterization of the stable maintenance properties of the par region of broad-host-range plasmid RK2. *J. Bacteriol.* **178**, 2086–2093 (1996).
71. Gerlitz, M., Hrabak, O. & Schwab, H. Partitioning of broad-host-range plasmid RP4 is a complex system involving site-specific recombination. *J. Bacteriol.* **172**, 6194–6203 (1990).
72. Easter, C. L., Schwab, H. & Helinski, D. R. Role of the parCBA operon of the broad-host-range plasmid RK2 in stable plasmid maintenance. *J. Bacteriol.* **180**, 6023–6030 (1998).
73. Tee, K. L. *et al.* An efficient transformation method for the bioplastic-producing " Knallgas " bacterium Ralstonia eutropha H16. *Biotechnol. J.* **12**, 1–7 (2017).
74. Park, H. C., Lim, K. J., Park, J. S., Lee, Y. H. & Huh, T. L. High frequency transformation of Alcaligenes eutrophus producing poly-β-hydroxybutyric acid by electroporation. *Biotechnol. Tech.* **9**, 31–34 (1995).
75. Tu, Q. *et al.* Room temperature electrocompetent bacterial cells improve DNA transformation and recombineering efficiency. *Sci. Rep.* **6**, 1–8 (2016).
76. Quandt, J. & Hynes, M. F. Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. *Gene* **127**, 15–21 (1993).
77. Lenz, O., Schwartz, E., Dernedde, J., Eitinger, M. & Friedrich, B. The Alcaligenes eutrophus H16 hoxX gene participates in hydrogenase regulation. *J. Bacteriol.* **176**, 4385–4393 (1994).
78. Reyrat, J. M., Pelicic, V., Gicquel, B. & Rappuoli, R. Counterselectable markers: Untapped tools for bacterial genetics and pathogenesis. *Infect. Immun.* **66**, 4011–4017 (1998).
79. Slater, S. *et al.* Multiple β-ketothiolases mediate poly(β-hydroxyalkanoate) copolymer synthesis in Ralstonia eutropha. *J. Bacteriol.* **180**, 1979–1987 (1998).
80. Arikawa, H., Matsumoto, K. & Fujiki, T. Polyhydroxyalkanoate production from sucrose by Cupriavidus necator strains harboring csc genes from Escherichia coli W. *Appl. Microbiol. Biotechnol.* **101**, 7497–7507 (2017).
81. Jeffke, T. *et al.* Mutational analysis of the cbb operon (CO2 assimilation) promoter of Ralstonia eutropha. *J. Bacteriol.* **184**, 4374–4380 (1999).
82. Park, J. M., Jang, Y. S., Kim, T. Y. & Lee, S. Y. Development of a gene knockout system for Ralstonia eutropha H16 based on the broad-host-range vector expressing a mobile group II intron. *FEMS Microbiol. Lett.* **309**, 193–200 (2010).
83. Doberstein, C., Grote, J., Wübbeler, J. H. & Steinbüchel, A. Polythioester synthesis in Ralstonia eutropha H16: Novel insights into 3,3'-thiodipropionic acid and 3,3'-

dithiodipropionic acid catabolism. *J. Biotechnol.* **184**, 187–198 (2014).

84. Johnson, B. F. & Stanier, R. Y. Dissimilation of aromatic compounds by Alcaligenes eutrophus. *J. Bacteriol.* **107**, 468–475 (1971).
85. Pérez-Pantoja, D., De La Iglesia, R., Pieper, D. H. & González, B. Metabolic reconstruction of aromatic compounds degradation from the genome of the amazing pollutant-degrading bacterium Cupriavidus necator JMP134. *FEMS Microbiol. Rev.* **32**, 736–794 (2008).
86. Jakočiūnas, T., Pedersen, L. E., Lis, A. V., Jensen, M. K. & Keasling, J. D. CasPER, a method for directed evolution in genomic contexts using mutagenesis and CRISPR/Cas9. *Metab. Eng.* **48**, 288–296 (2018).
87. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science (80-.).* **351**, (2016).
88. Brigham, C. J. *et al.* Elucidation of β-oxidation pathways in Ralstonia eutropha H16 by examination of global gene expression. *J. Bacteriol.* **192**, 5454–5464 (2010).
89. Zhang, J. & Greasham, R. Chemically defined media for commercial fermentations. *Appl. Microbiol. Biotechnol.* **51**, 407–421 (1999).
90. Terrade, N., Noël, R., Couillaud, R. & de Orduña, R. M. A new chemically defined medium for wine lactic acid bacteria. *Food Res. Int.* **42**, 363–367 (2009).
91. Oliver, P., P. & Sinskey, A. J. Poly-β-hydroxybutyrate (PHB) biosynthesis in Alcaligenes eutrophus H16. *J. Biol. Cemistry* **264**, 15298–15303 (1989).
92. Sharma, P. K. *et al.* Global changes in the proteome of Cupriavidus necator H16 during poly-(3-hydroxybutyrate) synthesis from various biodiesel by-product substrates. *AMB Express* **6**, 36 (2016).
93. Kumar, V., Bhalla, A. & Rathore, A. S. Design of experiments applications in bioprocessing: Concepts and approach. *Biotechnol. Prog.* **30**, 86–99 (2014).
94. Franceschini, G. & Macchietto, S. Model-based design of experiments for parameter precision: State of the art. *Chem. Eng. Sci.* **63**, 4846–4872 (2008).
95. Mandenius, C. & Brundin, A. Review: Biocatalysis and bioreactor design. *Biotechnol. Prog.* **24**, 1191–1203 (2008).
96. Niedz, R. P. & Evens, T. J. Design of experiments (DOE)—history, concepts, and relevance to in vitro culture. *Vitr. Cell. Dev. Biol. - Plant* **52**, 547–562 (2016).
97. Xu, P., Rizzoni, E. A., Sul, S. Y. & Stephanopoulos, G. Improving metabolic pathway efficiency by statistical model-based multivariate regulatory metabolic engineering. *ACS Synth. Biol.* **6**, 148–158 (2017).
98. Patel, K., Mark, W. & Claes, G. Leveraging gene synthesis, advanced cloning techniques, and machine learning for metabolic pathway engineering kedar. *Metab. Eng. Bioprocess Commer.* 1–122 (2016). doi:10.1007/978-3-319-41966-4

99. Caschera, F. *et al.* Coping with complexity: Machine learning optimization of cell-free protein synthesis. *Biotechnol. Bioeng.* **108**, 2218–2228 (2011).
100. Gustafsson, C. *et al.* Engineering genes for predictable protein expression. *Protein Expr. Purif.* **83**, 37–46 (2012).
101. Brown, S. R. *et al.* Design of experiments methodology to build a multifactorial statistical model describing the metabolic interactions of alcohol dehydrogenase isozymes in the ethanol biosynthetic pathway of the yeast Saccharomyces cerevisiae. *ACS Synth. Biol.* **7**, 1676–1684 (2018).
102. Lin, H., Kim, T., Xiong, F. & Yang, X. Enhancing the production of Fc fusion protein in fed-batch fermentation of Pichia pastoris by design of experiments. *Biotechnol. Prog.* **23**, 621–625 (2007).
103. Govindarajan, S. *et al.* Mapping of amino acid substitutions conferring herbicide resistance in wheat glutathione transferase. *ACS Synth. Biol.* **4**, 221–227 (2015).
104. Zhang, G., Mills, D. A. & Block, D. E. Development of chemically defined media supporting high-cell-density growth of lactococci, enterococci, and streptococci. *Appl. Environ. Microbiol.* **75**, 1080–1087 (2009).
105. Schneebeli, R. & Egli, T. A defined, Glucose-limited mineral medium for the cultivation of listeria spp. *Appl. Environ. Microbiol.* **79**, 2503–2511 (2013).
106. Sichwart, S., Hetzler, S., Bro, D. & Steinbu, A. Extension of the substrate utilization range of Ralstonia eutropha strain H16 by metabolic engineering to include mannose and glucose. *Appl. Environ. Microbiol.* **77**, 1325–1334 (2011).
107. Franz, A., Rehner, R., Kienle, A. & Grammel, H. Rapid selection of glucose-utilizing variants of the polyhydroxyalkanoate producer Ralstonia eutropha H16 by incubation with high substrate levels. *Lett. Appl. Microbiol.* **54**, 45–51 (2012).
108. Raberg, M. *et al.* Versatile metabolic adaptations of Ralstonia eutropha H16 to a loss of PdhL, the E3 component of the pyruvate dehydrogenase complex. *Appl. Environ. Microbiol.* **77**, 2254–2263 (2011).
109. Friedrich, C. G., Friedrich, B. & Bowien, B. Formation of enzymes of autotrophic metabolism during heterotrophic growth of Alcaligenes eutrophus. *Microbiol.* **122**, 69–78 (1981).
110. Alagesan, S., Minton, N. P. & Malys, N. 13C-assisted metabolic flux analysis to investigate heterotrophic and mixotrophic metabolism in Cupriavidus necator H16. *Metabolom.* **14**, (2018).
111. Shimizu, R. *et al.* New insight into the role of the calvin cycle: Reutilization of CO2 emitted through sugar degradation. *Sci. Rep.* **5**, (2015).
112. Shao, J., Marcondes, F. M. & Oliveira, V. Development of chemically defined media to express Trp-analog-labeled proteins in a Lactococcus lactis Trp auxotroph. *J. Mol.*

Microbiol. Biotechnol. **26**, 269–276 (2016).

113. Chandrangsu, P., Rensing, C. & Helmann, J. D. Metal homeostasis and resistance in bacteria. *Nat. Rev. Microbiol.* **15**, 338–350 (2017).
114. Argüello, J. M., Raimunda, D. & Padilla-benavides, T. Mechanisms of copper homeostasis in bacteria. *Front. Cell. Infect. Microbiol.* **3**, 1–14 (2013).
115. Festa, R. A. & Thiele, D. J. Copper: An essential metal in biology. *Curr. Biol.* **21**, 877–883 (2011).
116. Subramanyam, C. & Venkateswerlu, G. The effect of copper on histidine biosynthesis in Neurospora crassa. *J. Biosci.* **1**, 143–149 (1979).
117. Pearce, D. A. & Sherman, F. Toxicity of copper, cobalt, and nickel salts is dependent on histidine metabolism in the yeast Saccharomyces cerevisiae. *J. Bacteriol.* **181**, 4774–4779 (1999).
118. Müller, T., Walter, B., Wirtz, A. & Burkovski, A. Ammonium toxicity in bacteria. *Curr. Microbiol.* **52**, 400–406 (2006).
119. Meier-wagner, J., Nolden, L., Jakoby, M., Siewe, R. & Burkovski, A. Multiplicity of ammonium uptake systems in Corynebacterium glutamicum: role of Amt and AmtB. *Microbiol.* **147**, 135–143 (2017).
120. Detsch, C. & Stülke, J. Ammonium utilization in Bacillus subtilis: transport and regulatory functions of NrgA and NrgB. *Microbiol.* **149**, 3289–3297 (2017).
121. Dominguez, D. C. Calcium signalling in bacteria. *Mol. Microbiol.* **54**, 291–297 (2004).
122. Onoda, T., Enokizono, J., Kaya, H., Oshima, A. & Freestone, P. Effects of calcium and calcium chelators on growth and morphology of Escherichia coli L-form NC-7. *J. Bacteriol.* **182**, 1419–1422 (2000).
123. Herbaud, M. L., Guiseppi, A., Denizot, F., Haiech, J. & Kilhoffer, M. C. Calcium signalling in Bacillus subtilis. *Biochim. Biophys. Acta - Mol. Cell Res.* **1448**, 212–226 (1998).
124. Verissimo, L. M. P. *et al.* Influence of fructose on the diffusion of potassium hydrogen phosphate in aqueous solutions at 25 ° C. *J. Chem. Thermodyn.* **101**, 245–250 (2016).
125. Radchenko, M. V. *et al.* Potassium/proton antiport system of Escherichia coli. *J. Biol. Chem.* **281**, 19822–19829 (2006).
126. Ochrombel, I., Ott, L., Krämer, R., Burkovski, A. & Marin, K. Impact of improved potassium accumulation on pH homeostasis, membrane potential adjustment and survival of Corynebacterium glutamicum. *BBA - Bioenerg.* **1807**, 444–450 (2011).
127. Poolman, B. & Glaasker, E. Regulation of compatible solute accumulation in bacteria. *Mol. Microbiol.* **29**, 397–407 (1998).
128. Moncrief, M. B. C. & Maguire, M. E. Magnesium transport in prokaryotes. *J. Bio. Inorg. Chem.* **4**, 523–527 (1999).

129. Hartwig, A. Role of magnesium in genomic stability. *Mut. Res.* **475**, 113–121 (2001).
130. Smith, R. L. & Maguire, M. E. Microbial magnesium transport: Unusual transporters searching for identity. *Molecular Microbiology* **28**, 217–226 (1998).
131. Webb, M. The utilization of magnesium by certain Gram-Positive and Gram-negative bacteria. *J. Gen. Microbiol.* **43**, 401–409 (1966).
132. Nora, C., Martins-santana, L., Maria, L. & Monteiro, O. The art of vector engineering : towards the construction of next-generation genetic tools. *Microb. Biotechnol.* **12**, 125–147 (2019).
133. Martínez-García, E., Benedetti, I., Hueso, A. & De Lorenzo, V. Mining environmental plasmids for synthetic biology parts and devices. *Microbiol. Spectr.* **3**, 1–16 (2015).
134. Lee, M. E., C, D. W., Bernardo, C. & E, D. J. A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.* **4**, 975–986 (2015).
135. Bagdasarian, M. *et al.* Specific-purpose plasmid cloning vectors II. Broad host range, high copy number, RSF1010-derived gene cloning in Pseudomonas. *Gene* **16**, 237–247 (1981).
136. Taghavi, S., Van der Lelie, D. & Mergeay, M. Electroporation of Alcaligenes eutrophus with (Mega) plasmids and genomic DNA fragments. *Appl. Environ. Microbiol.* **60**, 3585–3591 (1994).
137. Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.* **166**, 557–580 (1983).
138. EUCAST. Determination of minimum inhibitory concentrations (MICs) of antibacterial agents by broth dilution: European Committee for Antimicrobial Susceptibility Testing (EUCAST) of the European Society of Clinical Microbiology and Infectious Disease. *Clin. Microbiol. Infect.* **9**, (2003).
139. Lee, H., Jeon, B. & Oh, M. Microbial production of ethanol from acetate by engineered Ralstonia eutropha. *Biotechnol. Biopro. Eng.* **21**, 402–407 (2016).
140. Siregar, J. J., Miroshnikov, K. & Mobashery, S. Purification, characterization, and investigation of the mechanism of aminoglycoside 3'-phosphotransferase type Ia. *Biochem.* **34**, 12681–12688 (1995).
141. Beck, E., Ludwig, G., Aucxswald, E. A., Reiss, B. & Schaller, H. Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from Tn5. *Gene* **19**, 327–336 (1982).
142. Wright, G. D. & Thompson, P. R. Aminoglycoside phosphotransferases: proteins, structure, and mechanism. *Front. Biosci.* **4**, D9-21 (1999).
143. Heuer, H. *et al.* Gentamicin resistance genes in environmental bacteria: Prevalence and transfer. *FEMS Microbiol. Ecol.* **42**, 289–302 (2002).
144. Díaz, M. A., Cooper, R. K., Cloeckaert, A. & Siebeling, R. J. Plasmid-mediated high-

level gentamicin resistance among enteric bacteria isolated from pet turtles in Louisiana. *Appl. Environ. Microbiol.* **72**, 306–312 (2006).

145. Kehrenberg, C., Catry, B., Haesebrouck, F., Kruif, A. De & Schwarz, S. Novel spectinomycin / streptomycin resistance gene, aadA14, from Pasteurella multocida. *Antimicrob. Agents Chemother.* **49**, 3046–3049 (2005).
146. Kotra, L. P., Haddad, J. & Mobashery, S. Aminoglycosides : perspectives on mechanisms of action and resistance and strategies to counter resistance. *Antimicrob. Agents Chemother.* **44**, 3249–3256 (2000).
147. Florentino, L. A. *et al.* Physiological and symbiotic diversity of Cupriavidus necator strains isolated from nodules of Leguminosae species. *Sci. Agric.* **69**, 247–258 (2012).
148. Roberts, C. Tetracycline resistance determinants : mechanisms of action, regulation of expression, genetic mobility, and distribution. *FEMS Microbiol. Rev.* **19**, 1–24 (1996).
149. Grossman, T. H. Tetracycline antibiotics and resistance. *Cold Spring Harb. Perspect. Med.* **6**, a025387 (2016).
150. Li, W. *et al.* Mechanism of tetracycline ribosomal protection protein Tet (O). *Nat. Comm.* **4**, 1477 (2013).
151. Amato, R. F. D., Thornsberry, C., Baker, C. N. & Kirven, L. A. Effect of calcium and magnesium ions on the susceptibility of Pseudomonas species to tetracycline, gentamicin polymyxin B, and carbenicillin. *Antimicrob. Agents Chemother.* **7**, 596–600 (1976).
152. Poehlsgaard, J. & Douthwaite, S. The bacterial ribosome as a target for antibiotics. *Nat. Rev. Microbiol.* **3**, 870–881 (2005).
153. Fernández, M. *et al.* Mechanisms of resistance to chloramphenicol in Pseudomonas putida KT2440. *Antimicrob. Agents Chemother.* **56**, 1001–1009 (2012).
154. Butler, K., English, A. R. & Ray, V. A. Carbenicillin : chemistry and mode of action. *J. Infect. Dis.* **122**, (2019).
155. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR / O, the TetR / O and AraC / I1 -I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
156. Knight, T. Idempotent vector design for standard assembly of biobricks standard biobrick sequence interface. *MIT Art. Intel. Lab.* 1–11 (2003).
157. Martínez-garcía, E., Calles, B., Arévalo-rodríguez, M. & Lorenzo, V. De. pBAM1 : an all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. *BMC Microbiol.* **11**, 1–13 (2011).
158. Silva-Rocha, R. *et al.* The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.* **41**, 666–675 (2013).

159. Shetty, R. P., Endy, D. & Jr, T. F. K. Engineering BioBrick vectors from BioBrick parts. *J. Bio. Eng.* **2**, 1–12 (2008).
160. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
161. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647 (2008).
162. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling : A one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One* **4**, e5553 (2009).
163. Anderson, J. C. *et al.* BglBricks : A flexible standard for biological part assembly. *J. Bio. Eng.* **4**, 1–12 (2010).
164. Jeong, J. *et al.* One-step sequence- and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. *Appl. Environ. Microbiol.* **78**, 5440–5443 (2012).
165. Trubitsyna, M., Michlewski, G., Cai, Y., Elfick, A. & French, E. PaperClip : rapid multi-part DNA assembly from existing libraries. *Nucleic Acids Res.* **42**, e154 (2014).
166. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6**, e16765 (2011).
167. Iverson, S. V, Haddock, T. L., Beal, J. & Densmore, D. M. CIDAR MoClo: Improved MoClo assembly standard and new E. coli part library enable rapid combinatorial design for synthetic and traditional biology. *ACS Synth. Biol.* **5**, 99–103 (2016).
168. Van Dolleweerd, C. J. *et al.* MIDAS: A modular DNA assembly system for synthetic biology. *ACS Synth. Biol.* **7**, 1018–1029 (2018).
169. Solaiman, D. K. Y. & Swingle, B. M. Isolation of novel Pseudomonas syringae promoters and functional characterization in polyhydroxyalkanoate-producing pseudomonads. *N. Biotechnol.* **27**, 1–9 (2010).
170. Martínez-Garćía, E., Aparicio, T., Goñi-Moreno, A., Fraile, S. & De Lorenzo, V. SEVA 2.0: An update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Res.* **43**, D1183–D1189 (2015).
171. Norrander, J., Kempe, T. & Messing, J. Construction of improved Ml3 vectors using oligodeoxynucleotidedirected mutagenesis (Ml3 cloning; synthetic DNA primers; gene synthesis; phosphoramidites; DNA polymerase I Klenow frag-ment). *Gene* **26**, 101–106 (1983).
172. Bolivar, F. *et al.* Constrcution and characterisation of new cloning vehicles. *Gene* **2**, 95–113 (1977).
173. Wang, B., Kitney, R. I., Joly, N. & Buck, M. Engineering modular and orthogonal

genetic logic gates for robust digital-like synthetic biology. *Nat. Commun.* **2**, 1–9 (2011).

174. Wei, H. L., Chakravarthy, S., Worley, J. N. & Collmer, A. Consequences of flagellin export through the type III secretion system of Pseudomonas syringae reveal a major difference in the innate immune systems of mammals and the model plant Nicotiana benthamiana. *Cell. Microbiol.* **15**, 601–618 (2013).
175. Sullivan, D. J. O. & Klaenhammer, T. R. High- and low-copy-number features for clone screening. *Gene* **137**, 227–231 (1993).
176. Haugan, K., Karunakaran, P., Tondervik, A. & Valla, S. The host range of RK2 minimal replicon copy-up mutants is limited by species-specific differences in the maximum tolerable copy number. *Plasmid* **33**, 27–39 (1995).
177. del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M. & Díaz-Orejas, R. Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* **62**, 434–64 (1998).
178. Chattoraj, D. K. Control of plasmid DNA replication by iterons: no longer paradoxical. *Mol. Microbiol.* **37**, 467–476 (2002).
179. Lee, J. W. *et al.* Resource creating single-copy genetic circuits. *Mol. Cell* **63**, 329–336 (2016).
180. Chang, A. C. Y. & Cohen, S. N. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* **134**, 1141–1156 (1978).
181. Kuest, U. & Stahl, U. L. F. Replication of plasmids in Gram-negative bacteria. *Microbiol. Rev.* **53**, 491–516 (1989).
182. Jiang, Y., Pacek, M., Helinski, D. R., Konieczny, I. & Toukdarian, A. A multifunctional plasmid-encoded replication initiation protein both recruits and positions an active helicase at the replication origin. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 8692–8697 (2003).
183. del Solar, G., Moscoso, M. & Espinosa, M. Rolling circle-replicating plasmids from Gram-positive and Gram-negative bacteria: a wall falls. *Mol. Microbiol.* **8**, 789–796 (1993).
184. Su, Y., Zhang, P. & Su, Y. An overview of biofuels policies and industrialization in the major biofuel producing countries. *Renew. Sustain. energy Rev.* **50**, 991–1003 (2015).
185. Peters, J. & Thielmann, S. Promoting biofuels: Implications for developing countries. *Energy Policy* **36**, 1538–1544 (2008).
186. Ajanovic, A. & Haas, R. Economic challenges for the future relevance of biofuels in transport in EU countries. *Energy* **35**, 3340–3348 (2010).
187. Demirbas, A. Political, economic and environmental impacts of biofuels: A review.

Appl. Energy **86**, 108–117 (2009).

188. Naik, S. N., Goud, V. V., Rout, P. K. & Dalai, A. K. Production of first and second generation biofuels: A comprehensive review. *Renew. Sustain. Energy Rev.* **14**, 578–597 (2010).
189. Carriquiry, M. A., Du, X. & Timilsina, G. R. Second generation biofuels: Economics and policies. *Energy Policy* **39**, 4222–4234 (2011).
190. Kim, T. H. & Kim, T. H. Overview of technical barriers and implementation of cellulosic ethanol in the U.S. *Energy* **66**, 13–19 (2014).
191. Alam, F., Mobin, S. & Chowdhury, H. Third generation biofuel from Algae. *Procedia Eng.* **105**, 763–768 (2015).
192. Yüksel, F. & Yüksel, B. The use of ethanol-gasoline blend as a fuel in an SI engine. *Renew. Energy* **29**, 1181–1191 (2004).
193. Howard, T. P. *et al.* Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7636–7641 (2013).
194. Lee, S. Y., Kim, H. M. & Cheon, S. Metabolic engineering for the production of hydrocarbon fuels. *Curr. Opin. Biotechnol.* **33**, 15–22 (2015).
195. Schirmer, A., Rude, M. A., Li, X., Popova, E. & del Cardayre, S. B. Microbial biosynthesis of alkanes. *Science (80-.).* **329**, 559–562 (2010).
196. Zhang, F., Rodriguez, S. & Keasling, J. D. Metabolic engineering of microbial pathways for advanced biofuels production. *Curr. Opin. Biotechnol.* **22**, 775–783 (2011).
197. Dunlop, M. J. *et al.* Engineering microbial biofuel tolerance and export using efflux pumps. *Mol. Syst. Biol.* **7**, 1–7 (2011).
198. Segura, A. *et al.* Solvent tolerance in Gram-negative bacteria. *Curr. Opin. Biotechnol.* **23**, 415–421 (2012).
199. Fischer, C. R., Klein-Marcuschamer, D. & Stephanopoulos, G. Selection and optimization of microbial hosts for biofuels production. *Metab. Eng.* **10**, 295–304 (2008).
200. Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. & Stephanopoulos, G. Tolerance and Production. *Science (80-.).* **314**, 1565–1568 (2006).
201. Alper, H. & Stephanopoulos, G. Global transcription machinery engineering: A new approach for improving cellular phenotype. *Metab. Eng.* **9**, 258–267 (2007).
202. Tan, F. *et al.* Using global transcription machinery engineering (gTME) to improve ethanol tolerance of Zymomonas mobilis. *Microb. Cell Fact.* **15**, 1–9 (2016).
203. Zhang, H., Chong, H., Ching, C. B., Song, H. & Jiang, R. Engineering global transcription factor cyclic AMP receptor protein of Escherichia coli for improved 1-

butanol tolerance. *Appl. Microbiol. Biotechnol.* **94**, 1107–1117 (2012).

204. Gao, X. *et al.* Tailoring of global transcription sigma D factor by random mutagenesis to improve Escherichia coli tolerance towards low-pHs. *J. Biotechnol.* **224**, 55–63 (2016).
205. Basak, S. & Jiang, R. Enhancing E. coli tolerance towards oxidative stress via engineering its global regulator cAMP receptor protein (CRP). *PLoS One* **7**, e51179 (2012).
206. Zhao, H., Li, J., Han, B., Li, X. & Chen, J. Improvement of oxidative stress tolerance in Saccharomyces cerevisiae through global transcription machinery engineering. *J. Ind. Microbiol. Biotechnol.* **41**, 869–878 (2014).
207. Ma, Y. & Yu, H. Engineering of Rhodococcus cell catalysts for tolerance improvement by sigma factor mutation and active plasmid partition. *J. Ind. Microbiol. Biotechnol.* **39**, 1421–1430 (2012).
208. Zhang, F. *et al.* Significantly improved solvent tolerance of Escherichia coli by global transcription machinery engineering. *Microb. Cell Fact.* **14**, 1–11 (2015).
209. Lin, Z., Zhang, Y. & Wang, J. Engineering of transcriptional regulators enhances microbial stress tolerance. *Biotechnol. Adv.* **31**, 986–991 (2013).
210. Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. & Stephanopoulos, G. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science (80-.).* **314**, 1565–1568 (2006).
211. Birnboim, H. C. & Doly, J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513–1524 (1979).
212. Kohse-Höinghaus, K. *et al.* Biofuel combustion chemistry: From ethanol to biodiesel. *Angew. Chemie - Int. Ed.* **49**, 3572–3597 (2010).
213. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).
214. Turner, N. J. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **5**, 567–573 (2009).
215. Swint-Kruse, L. Using evolution to guide protein engineering: The devil IS in the details. *Biophys. J.* **111**, 10–18 (2016).
216. Hendry, J. I., Bandyopadhyay, A., Srinivasan, S., Pakrasi, H. B. & Maranas, C. D. Metabolic model guided strain design of cyanobacteria. *Curr. Opin. Biotechnol.* **64**, 17–23 (2020).
217. Reeve, B., Martinez-Klimova, E., De Jonghe, J., Leak, D. J. & Ellis, T. The Geobacillus plasmid set: A modular toolkit for thermophile engineering. *ACS Synth. Biol.* **5**, 1342–1347 (2016).
218. Mejias Carpio, I. E., Machado-Santelli, G., Kazumi Sakata, S., Ferreira Filho, S. S. &

Rodrigues, D. F. Copper removal using a heavy-metal resistant microbial consortium in a fixed-bed reactor. *Water Res.* **62**, 156–166 (2014).

219. Ochoa-Herrera, V., León, G., Banihani, Q., Field, J. A. & Sierra-Alvarez, R. Toxicity of copper(II) ions to microorganisms in biological wastewater treatment systems. *Sci. Total Environ.* **412**–**413**, 380–385 (2011).
220. Hou, W., Chen, X., Song, G., Wang, Q. & Chi Chang, C. Effects of copper and cadmium on heavy metal polluted waterbody restoration by duckweed (Lemna minor). *Plant Physiol. Biochem.* **45**, 62–69 (2007).
221. Stewart, L. J. *et al.* Handling of nutrient copper in the bacterial envelope. *Metallomics* **11**, 50–63 (2019).
222. Djoko, K. Y. *et al.* Interplay between tolerance mechanisms to copper and acid stress in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 6818–6823 (2017).
223. Purohit, R. *et al.* Cu+-specific CopB transporter: Revising p1B-type ATPase classification. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 2108–2113 (2018).
224. Watanabe, D. *et al.* Exogenous addition of histidine reduces copper availability in the yeast Saccharomyces cerevisiae. *Microb. Cell* **1**, 241–246 (2014).
225. Deschamps, P., Kulkarni, P. P., Gautam-Basak, M. & Sarkar, B. The saga of copper(II)-L-histidine. *Coord. Chem. Rev.* **249**, 895–909 (2005).
226. Irtelli, B., Petrucci, W. A. & Navari-Izzo, F. Nicotianamine and histidine/proline are, respectively, the most important copper chelators in xylem sap of Brassica carinata under conditions of copper deficiency and excess. *J. Exp. Bot.* **60**, 269–277 (2009).
227. Vita, N. *et al.* Bacterial cytosolic proteins with a high capacity for Cu(I) that protect against copper toxicity. *Sci. Rep.* **6**, 1–11 (2016).
228. Ladomersky, E. & Petris, M. J. Copper tolerance and virulence in bacteria. *Metallomics* **7**, 957–964 (2015).
229. Kim, J., Darlington, A., Salvador, M., Utrilla, J. & Jiménez, J. I. Trade-offs between gene expression, growth and phenotypic diversity in microbial populations. *Curr. Opin. Biotechnol.* **62**, 29–37 (2020).
230. Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science (80-.).* **330**, 70–74 (2010).
231. Xu, P. *et al.* Modular optimization of multi-gene pathways for fatty acids production in E. coli. *Nat. Commun.* **4**, 1–8 (2013).
232. Wu, J., Du, G., Zhou, J. & Chen, J. Metabolic engineering of Escherichia coli for (2S)-pinocembrin production from glucose by a modular metabolic strategy. *Metab. Eng.* **16**, 48–55 (2013).
233. Chen, B., Lee, D. Y. & Chang, M. W. Combinatorial metabolic engineering of Saccharomyces cerevisiae for terminal alkene production. *Metab. Eng.* **31**, 53–61

(2015).

234. Elowitz, M. B. & Leibier, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
235. Chen, B. *et al.* Synthetic biology toolkits and applications in Saccharomyces cerevisiae. *Biotechnol. Adv.* **36**, 1870–1881 (2018).
236. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339–342 (2000).
237. Millet, L. J., Vélez, J. M. & Michener, J. K. Genetic Selection for Small Molecule Production in Competitive Microfluidic Droplets. *ACS Synth. Biol.* **8**, 1737–1743 (2019).
238. Joseph, R. C., Kim, N. M. & Sandoval, N. R. Recent developments of the synthetic biology toolkit for Clostridium. *Front. Microbiol.* **9**, 1–13 (2018).
239. Hatti-Kaul, R., Nilsson, L. J., Zhang, B., Rehnberg, N. & Lundmark, S. Designing Biobased Recyclable Polymers for Plastics. *Trends Biotechnol.* 1–18 (2019). doi:10.1016/j.tibtech.2019.04.011
240. Lynch, S., Eckert, C., Yu, J., Gill, R. & Maness, P. C. Overcoming substrate limitations for improved production of ethylene in E. coli. *Biotechnol. Biofuels* **9**, 1–10 (2016).
241. Ishihara, K. *et al.* Overexpression and In Vitro reconstitution of the Ethylene-forming enzyme from Pseudomonas syringae. *J. Fermen. Bioeng.* **79**, 205–211 (1995).
242. Fukuda, H. *et al.* Molecular cloning in Escherichia coli, expression, and nucleotide sequence of the gene for the ethylene-forming enzyme of Pseudomonas syringae pv. phaseolicola PK2. *Biochem. Biophy. Res. Comm.* **188**, 826–832 (1992).
243. Fukuda, H. *et al.* Heterologous expression of the gene for the ethylene-forming enzyme from Pseudomonas syringae in the cyanobacterium Synechococcus. *Biotechnol. Lett.* **16**, 1–6 (1994).
244. Pirkov, I., Albers, E., Norbeck, J. & Larsson, C. Ethylene production by metabolic engineering of the yeast Saccharomyces cerevisiae. *Metab. Eng.* **10**, 276–280 (2008).
245. Mycroft, Z., Gomis, M., Mines, P., Law, P. & Bugg, T. D. H. Biocatalytic conversion of lignin to aromatic dicarboxylic acids in Rhodococcus jostii RHA1 by re-routing aromatic degradation pathways. *Green Chem.* **17**, 4974–4979 (2015).
246. Liu, Y. *et al.* Engineered ethanol-driven biosynthetic system for improving production of acetyl-CoA derived drugs in Crabtree-negative yeast. *Metab. Eng.* **54**, 275–284 (2019).

Appendix A

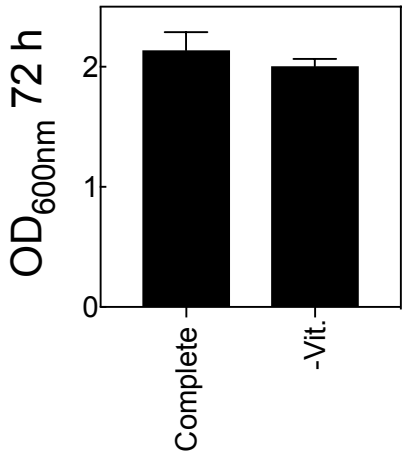


Fig. 1 Early withdrawal of vitamin solution from medium formulation. Growth was comparable to that of a complete medium formulation.

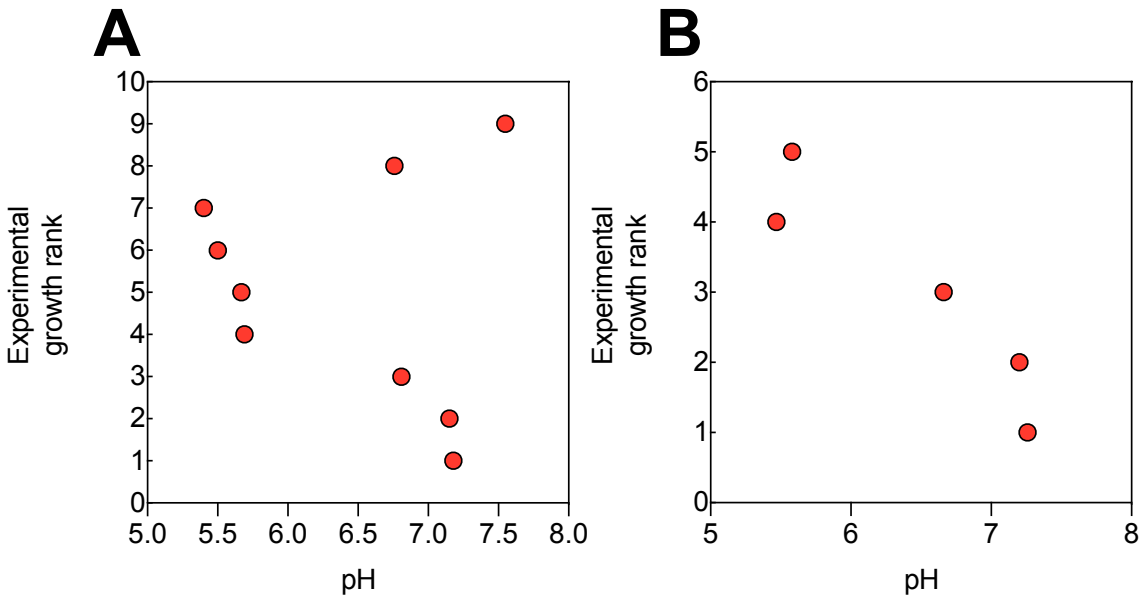


Fig. 2 Spearman's correlation of experimentally determined growth (OD_{600nm}) and pH after 72 h. **A.** Shake flask cultivation ($\rho = 0.23$). **B.** Bioreactor cultivation ($\rho = 0.52$). Spearman's correlation coefficient (ρ).

Appendix B

Reference to buffer for preparing chemically competent *C. necator*

Control buffer:
https://www.google.com.ng/url?sa=t&rct=j&q=&esrc=s&source=web&cd=11&cad=rja&uact=8&ved=2ahUKEwiZwfLvypPhAhXMRRUIHcWRBsIQFjAKegQIAhAC&url=http%3A%2F%2Fmcb.berkeley.edu%2Flabs%2Fkrantz%2Fprotocols%2Fcalcium_comp_cells.pdf&usg=AOvVaw12doCEp8nI668wL55RooLQ

Buffer B:
https://www.google.com.ng/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwiEv6vPzpPhAhVOShUIHSQDANEQFjAAegQIBRAB&url=https%3A%2F%2Fwww.neb.com%2Fprotocols%2F2012%2F06%2F21%2Fmaking-your-own-chemically-competent-cells&usg=AOvVaw3WbBLgLm_jQoOeUE97kbR8

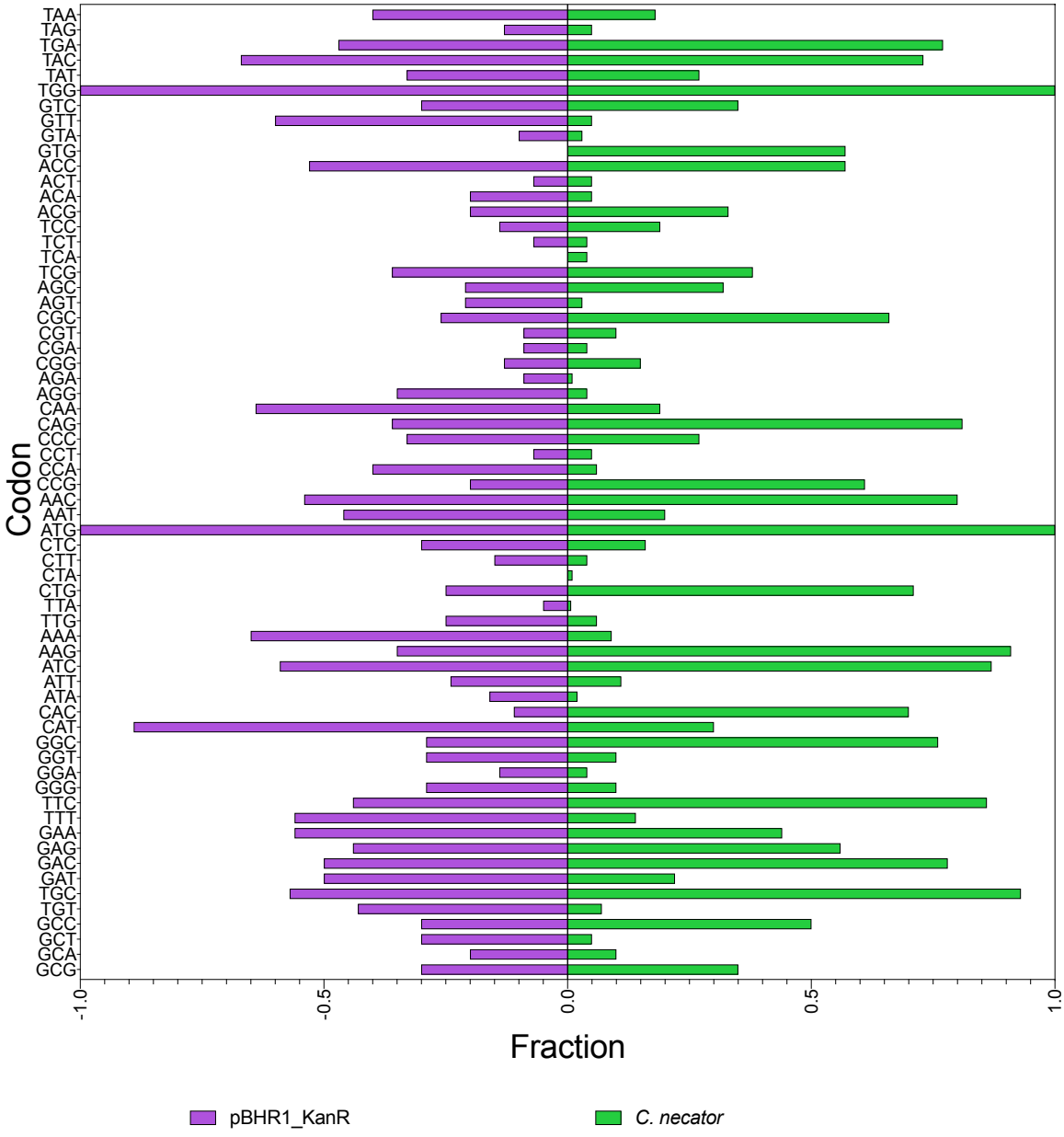
Buffer C:
https://www.google.com.ng/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjn9tP72JPhAhV0QRUIHbyND8EQFjAAegQIAhAB&url=http%3A%2F%2Fwww.molbi.de%2Fprotocols%2Fcompetent_cells_chemical_v1_0.htm&usg=AOvVaw3dOPIECieLe-DL_GHDBuSJ

Buffer D:
https://www.google.com.ng/url?sa=t&rct=j&q=&esrc=s&source=web&cd=19&cad=rja&uact=8&ved=2ahUKEwiVw7_i0pPhAhWhsXEKHc8CCMwQFjASegQIBxAC&url=https%3A%2F%2Fwww.helmholtz-muenchen.de%2Ffileadmin%2FPEPF%2FProtocols%2FChemically-competent-cells.pdf&usg=AOvVaw2bYy_1x2y0UNUV8bFpmwGN

Buffer E (RbCl was excluded):
https://www.google.com.ng/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwivvtaE55PhAhX5QRUIHZmZD8gQFjAAegQIBRAC&url=http%3A%2F%2Fwww.fechem.uzh.ch%2FMT%2Fpdf%2FComp_Cells.pdf&usg=AOvVaw13w6hpkwkwt_psO5cn0CFI

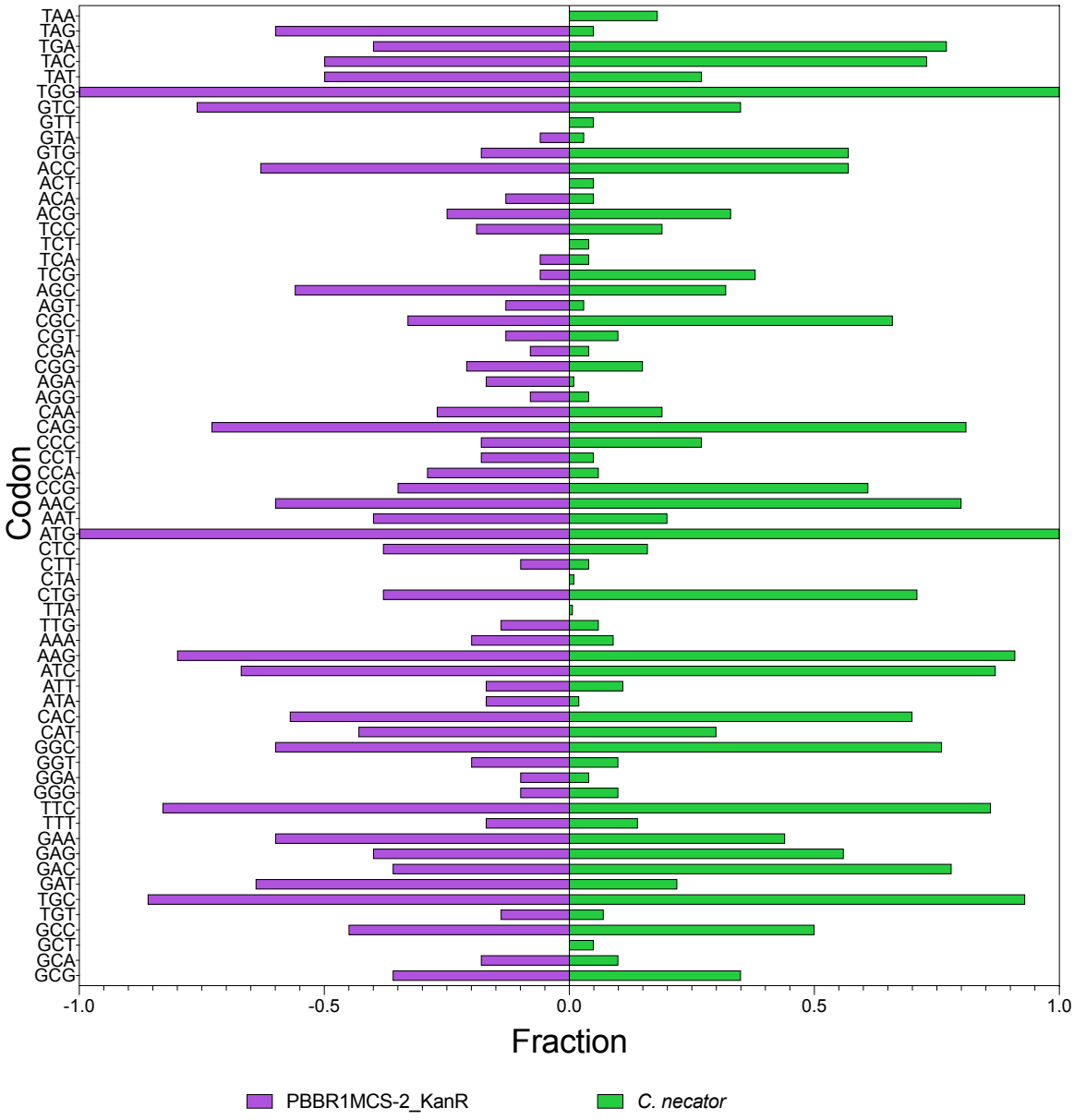
Buffer F:
https://www.google.com.ng/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjnrpLv75PhAhUFQhUIHZcwCs8QFjAAegQIARAB&url=https%3A%2F%2Fopenwetware.org%2Fwiki%2FPreparing_chemically_competent_cells&usg=AOvVaw1lY1cm4YDs159O_ekCX8Rz

Codon fraction of pBHR1 Kan sequence relative to *C. necator* codon fraction.



On the x-axis, the actual pBHR1_KanR fractions are positive values. *C. necator* codons table was retrieved from: https://www.kazusa.or.jp/codon/

Codon fraction of pBBR1MCS-2 Kan sequence relative to *C. necator* codon fraction.



On the x-axis, the actual pBBR1MCS-2_KanR fractions are positive values. *C. necator* codons table was retrieved from: https://www.kazusa.or.jp/codon/

Appendix C

Biological parts used for plasmid assembly

Three-part assembly

Part 1

***Cargoes*:**

PacI_MCS

GAACTTAATTAACTCATCGCAGTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAAC
GCGAATTTTAACAAAATATTAACGCTTACAATTTCCATTCGCCATTCAGGCTGCGCAACTGTTGGG
AAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAA
GGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGA
GCGCGCGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAG
AACTAGTGGATCCCCCGGGCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAG
GGGGGGCCCGGTACCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCA
TGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGA
AGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAAGC

PacI_P_{j5}_mRFP1_rrnB_T7

GAACTTAATTAAAGCGGATATAAAAACCGTTATTGACACAGGTGGAAATTTAGAATAACT
TGTTAGTAAACCTAATGGATCGACCTTATGGCGAGTAGCGAAGACGTTATCAAAGAGTT
CATGCGTTTCAAAGTTCGTATGGAAGGTTCCGTTAACGGTCACGAGTTCGAAATCGAAG
GTGAAGGTGAAGGTCGTCCGTACGAAGGTACCCAGACCGCTAAACTGAAAGTTACCAA
AGGTGGTCCGCTGCCGTTCGCTTGGGACATCCTGTCCCCGCAGTTCCAGTACGGTTCC
AAAGCTTACGTTAAACACCCGGCTGACATCCCGGACTACCTGAAACTGTCCTTCCCGGA
AGGTTTCAAATGGGAACGTGTTATGAACTTCGAAGACGGTGGTGTTGTTACCGTTACCC
AGGACTCCTCCCTGCAAGACGGTGAGTTCATCTACAAAGTTAAACTGCGTGGTACCAAC
TTCCCGTCCGACGGTCCGGTTATGCAGAAAAAAACCATGGGTTGGGAAGCTTCCACCG
AACGTATGTACCCGGAAGACGGTGCTCTGAAAGGTGAAATCAAAATGCGTCTGAAACTG
AAAGACGGTGGTCACTACGACGCTGAAGTTAAAACCACCTACATGGCTAAAAAACCGGT
TCAGCTGCCGGGTGCTTACAAAACCGACATCAAACTGGACATCACCTCCCACAACGAAG
ACTACACCATCGTTGAACAGTACGAACGTGCTGAAGGTCGTCACTCCACCGGTGCTTAA
GGATCCAAACTCGAGTAAGGATCTCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAA
GACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACT
GGCTCACCTTCGGGTGGGCCTTTCTGCGAAGC

PacI_$P_{j5[C2]}$_mRFP1_rrnB_T7

GAACTTAATTAAAGCGGATATAAAAACCGTTATTGACACAGGTGGAAATTTAGAATAACT
TGTTAGTAAACCTAATGGATCGACCTTAGATCTTTTAAGAAGGAGATATACATATGGCGA
GTAGCGAAGACGTTATCAAAGAGTTCATGCGTTTCAAAGTTCGTATGGAAGGTTCCGTT
AACGGTCACGAGTTCGAAATCGAAGGTGAAGGTGAAGGTCGTCCGTACGAAGGTACCC
AGACCGCTAAACTGAAAGTTACCAAAGGTGGTCCGCTGCCGTTCGCTTGGGACATCCT
GTCCCCGCAGTTCCAGTACGGTTCCAAAGCTTACGTTAAACACCCGGCTGACATCCCG
GACTACCTGAAACTGTCCTTCCCGGAAGGTTTCAAATGGGAACGTGTTATGAACTTCGA
AGACGGTGGTGTTGTTACCGTTACCCAGGACTCCTCCCTGCAAGACGGTGAGTTCATCT
ACAAAGTTAAACTGCGTGGTACCAACTTCCCGTCCGACGGTCCGGTTATGCAGAAAAAA
ACCATGGGTTGGGAAGCTTCCACCGAACGTATGTACCCGGAAGACGGTGCTCTGAAAG
GTGAAATCAAAATGCGTCTGAAACTGAAAGACGGTGGTCACTACGACGCTGAAGTTAAA
ACCACCTACATGGCTAAAAAACCGGTTCAGCTGCCGGGTGCTTACAAAACCGACATCAA
ACTGGACATCACCTCCCACAACGAAGACTACACCATCGTTGAACAGTACGAACGTGCTG
AAGGTCGTCACTCCACCGGTGCTTAAGGATCCAAACTCGAGTAAGGATCTCCAGGCATC
AAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGG
TGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGAAGC

PacI_P$_{j5[C2]}$_eGFP_rrnB_T7

GAACTTAATTAAAGCGGATATAAAAACCGTTATTGACACAGGTGGAAATTTAGAATAACT
TGTTAGTAAACCTAATGGATCGACCTTAGATCTTTTAAGAAGGAGATATACATATGCGTA
AAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTA
ATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTT
ACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACT
ACTTTCGGTTATGGTGTTCAATGCTTTGCGAGATACCCAGATCATATGAAACAGCATGAC
TTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGAT
GACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAG
AATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGA
ATACAACTATAACTCACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCAA
AGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTA
TCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTC
CACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGATCACATGGTCCTTCTTG
AGTTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAACCAGGC
ATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGT
CGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGA
AGC

PacI_P$_{tac[C2]}$_mRFP1_rrnB_T7

GAACTTAATTAATTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGATCTTT
TAAGAAGGAGATATACATATGGCGAGTAGCGAAGACGTTATCAAAGAGTTCATGCGTTT
CAAAGTTCGTATGGAAGGTTCCGTTAACGGTCACGAGTTCGAAATCGAAGGTGAAGGTG
AAGGTCGTCCGTACGAAGGTACCCAGACCGCTAAACTGAAAGTTACCAAAGGTGGTCC
GCTGCCGTTCGCTTGGGACATCCTGTCCCCGCAGTTCCAGTACGGTTCCAAAGCTTAC
GTTAAACACCCGGCTGACATCCCGGACTACCTGAAACTGTCCTTCCCGGAAGGTTTCAA
ATGGGAACGTGTTATGAACTTCGAAGACGGTGGTGTTGTTACCGTTACCCAGGACTCCT
CCCTGCAAGACGGTGAGTTCATCTACAAAGTTAAACTGCGTGGTACCAACTTCCCGTCC
GACGGTCCGGTTATGCAGAAAAAAACCATGGGTTGGGAAGCTTCCACCGAACGTATGTA
CCCGGAAGACGGTGCTCTGAAAGGTGAAATCAAAATGCGTCTGAAACTGAAAGACGGT
GGTCACTACGACGCTGAAGTTAAAACCACCTACATGGCTAAAAAACCGGTTCAGCTGCC
GGGTGCTTACAAAACCGACATCAAACTGGACATCACCTCCCACAACGAAGACTACACCA
TCGTTGAACAGTACGAACGTGCTGAAGGTCGTCACTCCACCGGTGCTTAAGGATCCAAA
CTCGAGTAAGGATCTCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGC
CTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACC
TTCGGGTGGGCCTTTCTGCGAAGC

PacI-P$_{tac_[C2]}$_eGFP-rrnB-T7

GAACTTAATTAATTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGATCTTT
TAAGAAGGAGATATACATATGCGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATT
CTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGA
AGGTGATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACC
TGTTCCATGGCCAACACTTGTCACTACTTTCGGTTATGGTGTTCAATGCTTTGCGAGATA
CCCAGATCATATGAAACAGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTAC
AGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAAGACACGTGCTGAAGTCAAG
TTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGAT
GGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATGTATACATCATG
GCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGAT
GGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGT
CCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGA

AAAGAGAGATCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCA
TGGATGAACTATACAAATAACCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACT
GGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCT
CACCTTCGGGTGGGCCTTTCTGCGAAGC

PacI-J$_{23100_PETRBS}$-mRFP1-BBa_B0015

GAACTTAATTAAGCCGCTTCTAGAACGTCTCAATCTCTATTTGACGGCTAGCTCAGTCCT
AGGTACAGTGCTAGCGTACTTTAACTTTAAGAAGGAGATATACATATGGCGAGTAGCGA
AGACGTTATCAAAGAGTTCATGCGTTTCAAAGTTCGTATGGAAGGTTCCGTTAACGGTCA
CGAGTTCGAAATCGAAGGTGAAGGTGAAGGTCGTCCGTACGAAGGTACCCAGACCGCT
AAACTGAAAGTTACCAAAGGTGGTCCGCTGCCGTTCGCTTGGGACATCCTGTCCCCGC
AGTTCCAGTACGGTTCCAAAGCTTACGTTAAACACCCGGCTGACATCCCGGACTACCTG
AAACTGTCCTTCCCGGAAGGTTTCAAATGGGAACGTGTTATGAACTTCGAAGACGGTGG
TGTTGTTACCGTTACCCAGGACTCCTCCCTGCAAGACGGTGAGTTCATCTACAAAGTTA
AACTGCGTGGTACCAACTTCCCGTCCGACGGTCCGGTTATGCAGAAAAAAACCATGGGT
TGGGAAGCTTCCACCGAACGTATGTACCCGGAAGACGGTGCTCTGAAAGGTGAAATCA
AAATGCGTCTGAAACTGAAAGACGGTGGTCACTACGACGCTGAAGTTAAAACCACCTAC
ATGGCTAAAAAACCGGTTCAGCTGCCGGGTGCTTACAAAACCGACATCAAACTGGACAT
CACCTCCCACAACGAAGACTACACCATCGTTGAACAGTACGAACGTGCTGAAGGTCGTC
ACTCCACCGGTGCTTAAGGATCCAAACTCGAGTAAGGATCTCCAGGCATCAAATAAAAC
GAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCT
CTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTA AAGC

PacI_J$_{23100_PETRBS}$_eGFP-BBa_B0015

GAACTTAATTAAGCCGCTTCTAGAACGTCTCAATCTCTATTTGACGGCTAGCTCAGTCCT
AGGTACAGTGCTAGCGTACTTTAACTTTAAGAAGGAGATATACATATGCGTAAAGGAGAA
GAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCAC
AAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAA
ATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCGG
TTATGGTGTTCAATGCTTTGCGAGATACCCAGATCATATGAAACAGCATGACTTTTTCAA
GAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGA
ACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGT
TAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACT
ATAACTCACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACT
TCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAA
ATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAAT
CTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGATCACATGGTCCTTCTTGAGTTTGTAA
CAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAATCGACCAGGCATCAA
ATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTG
AACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTAAAG
C

Part 2

Resistant cassette:

AscI_CmR: pBHR1

AAGCGGCGCGCCTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTTTCG
AATTTCTGCCATTCATCCGCTTATTATACTTATTCAGGCGTAGCACCAGGCGTTTAAGGG
CACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTG
TAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAGACGGCATGATGAACCTGAA
TCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACG
GGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCA
GGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT
TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG
TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACA
AGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCG
GATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTA
TTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTA
CATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCA
ACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCG
ATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTC
TTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCA
ACAGGGACACCAGGT

AscI_KanR: pBHR1

AAGCGGCGCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCAATTAA
CCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAG
GATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGA
GGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACA
TCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCAT
GAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGT
TCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTC
ATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACA
AACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCAC
CTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTG
AGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAA
TTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTT
GCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCG
CACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGT
TGGAATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCC
CTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATCTTG
TGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCAGGT

AscI_TetR: pT18mobScab

AAGCGGCGCGCCGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAA
GGATCTGATGGCGCAGGGGATCAAGATCTCCGAAAAGTGCCACCTGACGTCTAAGAAA
CCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTC
AAGAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGT
TAAATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATC
CTCGGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGG
GCCTCTTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTA
GCGCTATATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCG
CTTTGGCCGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCG
ATCATGGCGACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCA

TCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGG
AAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGC
AGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCG
GCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGC
ATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCG
GTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAAC
TCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTG
GAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCT
CAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCG
CCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGACGCGAGGCT
GGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCCCGCGTT
GCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGATCG
CTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGATTTA
TGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATAC
CTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAA
TGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCATGGCGATGCCTGCTTGCC
GAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCAGGT

Part 3

Replication sequence:

FseI_*Rep*: pBHR1

AGGTGGCCGGCCATAATTGTTGTCGCGCTGCCGAAAAGTTGCAGCTGATTGCGCATGG
TGCCGCAACCGTGCGGCACCCTACCGCATGGAGATAAGCATGGCCACGCAGTCCAGA
GAAATCGGCATTCAAGCCAAGAACAAGCCCGGTCACTGGGTGCAAACGGAACGCAAAG
CGCATGAGGCGTGGGCCGGGCTTATTGCGAGGAAACCCACGGCGGCAATGCTGCTGC
ATCACCTCGTGGCGCAGATGGGCCACCAGAACGCCGTGGTGGTCAGCCAGAAGACACT
TTCCAAGCTCATCGGACGTTCTTTGCGGACGGTCCAATACGCAGTCAAGGACTTGGTGG
CCGAGCGCTGGATCTCCGTCGTGAAGCTCAACGGCCCCGGCACCGTGTCGGCCTACG
TGGTCAATGACCGCGTGGCGTGGGGCCAGCCCCGCGACCAGTTGCGCCTGTCGGTGT
TCAGTGCCGCCGTGGTGGTTGATCACGACGACCAGGACGAATCGCTGTTGGGGCATGG
CGACCTGCGCCGCATCCCGACCCTGTATCCGGGCGAGCAGCAACTACCGACCGGCCC
CGGCGAGGAGCCGCCCAGCCAGCCCGGCATTCCGGGCATGGAACCAGACCTGCCAGC
CTTGACCGAAACGGAGGAATGGGAACGGCGCGGGCAGCAGCGCCTGCCGATGCCCGA
TGAGCCGTGTTTTCTGGACGATGGCGAGCCGTTGGAGCCGCCGACACGGGTCACGCT
GCCGCGCCGGTAGCACTTGGGTTGCGCAGCAACCCGTAAGTGCGCTGTTCCAGACTAT
CGGCTGTAGCCGCCTCGCCGCCCTATACCTTGTCTGGAAC

FseI_*Oriv-Rep*: pBBR1MCS-2

AGGTGGCCGGCCCTTCGCAAAGTCGTGACCGCCTACGGCGGCTGCGGCGCCCTACGG
GCTTGCTCTCCGGGCTTCGCCCTGCGCGGTCGCTGCGCTCCCTTGCCAGCCCGTGGAT
ATGTGGACGATGGCCGCGAGCGGCCACCGGCTGGCTCGCTTCGCTCGGCCCGTGGAC
AACCCTGCTGGACAAGCTGATGGACAGGCTGCGCCTGCCCACGAGCTTGACCACAGG
GATTGCCCACCGGCTACCCAGCCTTCGACCACATACCCACCGGCTCCAACTGCGCGGC
CTGCGGCCTTGCCCCATCAATTTTTTTAATTTTCTCTGGGGAAAAGCCTCCGGCCTGCG
GCCTGCGCGCTTCGCTTGCCGGTTGGACACCAAGTGGAAGGCGGGTCAAGGCTCGCG
CAGCGACCGCGCAGCGGCTTGGCCTTGACGCGCCTGGAACGACCCAAGCCTATGCGA
GTGGGGGCAGTCGAAGGCGAAGCCCGCCCGCCTGCCCCCCGAGCCTCACGGCGGCG
AGTGCGGGGGTTCCAAGGGGGCAGCGCCACCTTGGGCAAGGCCGAAGGCCGCGCAG
TCGATCAACAAGCCCCGGAGGGGCCACTTTTTGCCGGAGGGGGAGCCGCGCCGAAGG
CGTGGGGGAACCCCGCAGGGGTGCCCTTCTTTGGGCACCAAAGAACTAGATATAGGGC

GAAATGCGAAAGACTTAAAAATCAACAACTTAAAAAAGGGGGGTACGCAACAGCTCATT
GCGGCACCCCCCGCAATAGCTCATTGCGTAGGTTAAAGAAAATCTGTAATTGACTGCCA
CTTTTACGCAACGCATAATTGTTGTCGCGCTGCCGAAAAGTTGCAGCTGATTGCGCATG
GTGCCGCAACCGTGCGGCACCCTACCGCATGGAGATAAGCATGGCCACGCAGTCCAG
AGAAATCGGCATTCAAGCCAAGAACAAGCCCGGTCACTGGGTGCAAACGGAACGCAAA
GCGCATGAGGCGTGGGCCGGGCTTATTGCGAGGAAACCCACGGCGGCAATGCTGCTG
CATCACCTCGTGGCGCAGATGGGCCACCAGAACGCCGTGGTGGTCAGCCAGAAGACA
CTTTCCAAGCTCATCGGACGTTCTTTGCGGACGGTCCAATACGCAGTCAAGGACTTGGT
GGCCGAGCGCTGGATCTCCGTCGTGAAGCTCAACGGCCCCGGCACCGTGTCGGCCTA
CGTGGTCAATGACCGCGTGGCGTGGGGCCAGCCCCGCGACCAGTTGCGCCTGTCGGT
GTTCAGTGCCGCCGTGGTGGTTGATCACGACGACCAGGACGAATCGCTGTTGGGGCAT
GGCGACCTGCGCCGCATCCCGACCCTGTATCCGGGCGAGCAGCAACTACCGACCGGC
CCCGGCGAGGAGCCGCCCAGCCAGCCCGGCATTCCGGGCATGGAACCAGACCTGCCA
GCCTTGACCGAAACGGAGGAATGGGAACGGCGCGGGCAGCAGCGCCTGCCGATGCCC
GATGAGCCGTGTTTTCTGGACGATGGCGAGCCGTTGGAGCCGCCGACACGGGTCACG
CTGCCGCGCCGGTAGCACTTGGGTTGCGCAGCAACCCGTAAGTGCGCTGTTCCAGACT
ATCGGCTGTAGCCGCCTCGCCGCCCTATACCTTGTCTGGAAC

FseI_*OriV*: pBBR1MCS-2

AGGTGGCCGGCCCTTCGCAAAGTCGTGACCGCCTACGGCGGCTGCGGCGCCCTACGG
GCTTGCTCTCCGGGCTTCGCCCTGCGCGGTCGCTGCGCTCCCTTGCCAGCCCGTGGAT
ATGTGGACGATGGCCGCGAGCGGCCACCGGCTGGCTCGCTTCGCTCGGCCCGTGGAC
AACCCTGCTGGACAAGCTGATGGACAGGCTGCGCCTGCCCACGAGCTTGACCACAGG
GATTGCCCACCGGCTACCCAGCCTTCGACCACATACCCACCGGCTCCAACTGCGCGGC
CTGCGGCCTTGCCCCATCAATTTTTTTAATTTTCTCTGGGGAAAAGCCTCCGGCCTGCG
GCCTGCGCGCTTCGCTTGCCGGTTGGACACCAAGTGGAAGGCGGGTCAAGGCTCGCG
CAGCGACCGCGCAGCGGCTTGGCCTTGACGCGCCTGGAACGACCCAAGCCTATGCGA
GTGGGGGCAGTCGAAGGCGAAGCCCGCCCGCCTGCCCCCCGAGCCTCACGGCGGCG
AGTGCGGGGGTTCCAAGGGGGCAGCGCCACCTTGGGCAAGGCCGAAGGCCGCGCAG
TCGATCAACAAGCCCCGGAGGGGCCACTTTTTGCCGGAGGGGGAGCCGCGCCGAAGG
CGTGGGGGAACCCCGCAGGGGTGCCCTTCTTTGGGCACCAAAGAACTAGATATAGGGC
GAAATGCGAAAGACTTAAAAATCAACAACTTAAAAAAGGGGGGTACGCAACAGCTCATT
GCGGCACCCCCCGCAATAGCTCATTGCGTAGGTTAAAGAAAATCTGTAATTGACTGCCA
CTTTTACGCAACGCATAATTGTTGTCGCGCTGCCGAAAAGTTGCAGCTGATTGCGCATG
GTGCCGCAACCGTGCGGCACCCTACCGCATGGAGATAAGCGAAC

FseI_*Rep*: pBBR1MCS-2

AGGTGGCCGGCCATGGCCACGCAGTCCAGAGAAATCGGCATTCAAGCCAAGAACAAGC
CCGGTCACTGGGTGCAAACGGAACGCAAAGCGCATGAGGCGTGGGCCGGGCTTATTG
CGAGGAAACCCACGGCGGCAATGCTGCTGCATCACCTCGTGGCGCAGATGGGCCACC
AGAACGCCGTGGTGGTCAGCCAGAAGACACTTTCCAAGCTCATCGGACGTTCTTTGCG
GACGGTCCAATACGCAGTCAAGGACTTGGTGGCCGAGCGCTGGATCTCCGTCGTGAAG
CTCAACGGCCCCGGCACCGTGTCGGCCTACGTGGTCAATGACCGCGTGGCGTGGGGC
CAGCCCCGCGACCAGTTGCGCCTGTCGGTGTTCAGTGCCGCCGTGGTGGTTGATCACG
ACGACCAGGACGAATCGCTGTTGGGGCATGGCGACCTGCGCCGCATCCCGACCCTGT
ATCCGGGCGAGCAGCAACTACCGACCGGCCCCGGCGAGGAGCCGCCCAGCCAGCCC
GGCATTCCGGGCATGGAACCAGACCTGCCAGCCTTGACCGAAACGGAGGAATGGGAA
CGGCGCGGGCAGCAGCGCCTGCCGATGCCCGATGAGCCGTGTTTTCTGGACGATGGC
GAGCCGTTGGAGCCGCCGACACGGGTCACGCTGCCGCGCCGGTAGCACTTGGGTTGC
GCAGCAACCCGTAAGTGCGCTGTTCCAGACTATCGGCTGTAGCCGCCTCGCCGCCCTA
TACCTTGTCTGGAAC

FseI_pMB1: pBR322 (*Rop_ori*)

AGGTGGCCGGCCTCCAGTAACCGGGCATGTTCATCATCAGTAACCCGTATCGTGAGCA
TCCTCTCTCGTTTCATCGGTATCATTACCCCCATGAACAGAAATCCCCCTTACACGGAG
GCATCAGTGACCAAACAGGAAAAAACCGCCCTTAACATGGCCCGCTTTATCAGAAGCCA
GACATTAACGCTTCTGGAGAAACTCAACGAGCTGGACGCGGATGAACAGGCAGACATC
TGTGAATCGCTTCACGACCACGCTGATGAGCTTTACCGCAGCTGCCTCGCGCGTTTCG
GTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCT
GTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCG
GGTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTT
AACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACC
GCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACT
GACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCG
GTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGG
CCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTC
CGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGA
CAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTT
CCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGC
TTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTG
GGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATC
GTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAA
CAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCT
AACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC
CTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGT
GGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCT
TTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAGGAAC

FseI_ pMB1: pUC19 (*ori*)

AGGTGGCCGGCCGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGC
GTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCT
CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGG
AAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCC
TTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTC
GGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGAC
CGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATC
GCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCT
ACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTAT
CTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCA
AACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGA
AAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCGAAC

FseI_p15A: pBW115lac_hrpR

AGGTGGCCGGCCCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGAC
CAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTA
GGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCA
CTGAGCGTCAGACCCCTTAATAAGATGATCTTCTTGAGATCGTTTTGGTCTGCGCGTAAT
CTCTTGCTCTGAAAACGAAAAAACCGCCTTGCAGGGCGGTTTTTCGAAGGTTCTCTGAG
CTACCAACTCTTTGAACCGAGGTAACTGGCTTGGAGGAGCGCAGTCACCAAAACTTGTC
CTTTCAGTTTAGCCTTAACCGGCGCATGACTTCAAGACTAACTCCTCTAAATCAATTACC
AGTGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTCAAGACGATAGT
TACCGGATAAGGCGCAGCGGTCGGACTGAACGGGGGGTTCGTGCATACAGTCCAGCTT
GGAGCGAACTGCCTACCCGGAACTGAGTGTCAGGCGTGGAATGAGACAAACGCGGCC

ATAACAGCGGAATGACACCGGTAAACCGAAAGGCAGGAACAGGAGAGCGCACGAGGG
AGCCGCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCACTG
ATTTGAGCGTCAGATTTCGTGATGCTTGTCAGGGGGGCGGAGCCTATGGAAAAACGGC
TTTGCCGCGGCCCTCTCACTTCCCTGTTAAGTATCTTCCTGGCATCTTCCAGGAAATCTC
CGCCCCGTTCGTAAGCCATTTCCGCGAAC

Four bio-parts assembly

Part 1

Cargoes:

Same as in the three bio-parts assembly

Part 2

Antibiotic cassette:

AscI_KanR from pBHR1

Same as the three bio-parts assembly

Part 3

Replication sequence:

FseI_*Rep*: pBHR1

AGGTGGCCGGCCATAATTGTTGTCGCGCTGCCGAAAAGTTGCAGCTGATTGCGCATGG
TGCCGCAACCGTGCGGCACCCTACCGCATGGAGATAAGCATGGCCACGCAGTCCAGA
GAAATCGGCATTCAAGCCAAGAACAAGCCCGGTCACTGGGTGCAAACGGAACGCAAAG
CGCATGAGGCGTGGGCCGGGCTTATTGCGAGGAAACCCACGGCGGCAATGCTGCTGC
ATCACCTCGTGGCGCAGATGGGCCACCAGAACGCCGTGGTGGTCAGCCAGAAGACACT
TTCCAAGCTCATCGGACGTTCTTTGCGGACGGTCCAATACGCAGTCAAGGACTTGGTGG
CCGAGCGCTGGATCTCCGTCGTGAAGCTCAACGGCCCCGGCACCGTGTCGGCCTACG
TGGTCAATGACCGCGTGGCGTGGGGCCAGCCCCGCGACCAGTTGCGCCTGTCGGTGT
TCAGTGCCGCCGTGGTGGTTGATCACGACGACCAGGACGAATCGCTGTTGGGGCATGG
CGACCTGCGCCGCATCCCGACCCTGTATCCGGGCGAGCAGCAACTACCGACCGGCCC
CGGCGAGGAGCCGCCCAGCCAGCCCGGCATTCCGGGCATGGAACCAGACCTGCCAGC
CTTGACCGAAACGGAGGAATGGGAACGGCGCGGGCAGCAGCGCCTGCCGATGCCCGA
TGAGCCGTGTTTTCTGGACGATGGCGAGCCGTTGGAGCCGCCGACACGGGTCACGCT
GCCGCGCCGGTAGCACTTGGGTTGCGCAGCAACCCGTAAGTGCGCTGTTCCAGACTAT
CGGCTGTAGCCGCCTCGCCGCCCTATACCTTGTCTGGGTA

FseI_*Oriv-Rep*: pBBR1MCS-2

AGGTGGCCGGCCCTTCGCAAAGTCGTGACCGCCTACGGCGGCTGCGGCGCCCTACGG
GCTTGCTCTCCGGGCTTCGCCCTGCGCGGTCGCTGCGCTCCCTTGCCAGCCCGTGGAT
ATGTGGACGATGGCCGCGAGCGGCCACCGGCTGGCTCGCTTCGCTCGGCCCGTGGAC
AACCCTGCTGGACAAGCTGATGGACAGGCTGCGCCTGCCCACGAGCTTGACCACAGG
GATTGCCCACCGGCTACCCAGCCTTCGACCACATACCCACCGGCTCCAACTGCGCGGC
CTGCGGCCTTGCCCCATCAATTTTTTTAATTTTCTCTGGGGAAAAGCCTCCGGCCTGCG
GCCTGCGCGCTTCGCTTGCCGGTTGGACACCAAGTGGAAGGCGGGTCAAGGCTCGCG
CAGCGACCGCGCAGCGGCTTGGCCTTGACGCGCCTGGAACGACCCAAGCCTATGCGA
GTGGGGGCAGTCGAAGGCGAAGCCCGCCCGCCTGCCCCCCGAGCCTCACGGCGGCG
AGTGCGGGGGTTCCAAGGGGGCAGCGCCACCTTGGGCAAGGCCGAAGGCCGCGCAG
TCGATCAACAAGCCCCGGAGGGGCCACTTTTTGCCGGAGGGGGAGCCGCGCCGAAGG

CGTGGGGGAACCCCGCAGGGGTGCCCTTCTTTGGGCACCAAAGAACTAGATATAGGGC
GAAATGCGAAAGACTTAAAAATCAACAACTTAAAAAAGGGGGGTACGCAACAGCTCATT
GCGGCACCCCCCGCAATAGCTCATTGCGTAGGTTAAAGAAAATCTGTAATTGACTGCCA
CTTTTACGCAACGCATAATTGTTGTCGCGCTGCCGAAAAGTTGCAGCTGATTGCGCATG
GTGCCGCAACCGTGCGGCACCCTACCGCATGGAGATAAGCATGGCCACGCAGTCCAG
AGAAATCGGCATTCAAGCCAAGAACAAGCCCGGTCACTGGGTGCAAACGGAACGCAAA
GCGCATGAGGCGTGGGCCGGGCTTATTGCGAGGAAACCCACGGCGGCAATGCTGCTG
CATCACCTCGTGGCGCAGATGGGCCACCAGAACGCCGTGGTGGTCAGCCAGAAGACA
CTTTCCAAGCTCATCGGACGTTCTTTGCGGACGGTCCAATACGCAGTCAAGGACTTGGT
GGCCGAGCGCTGGATCTCCGTCGTGAAGCTCAACGGCCCCGGCACCGTGTCGGCCTA
CGTGGTCAATGACCGCGTGGCGTGGGGCCAGCCCCGCGACCAGTTGCGCCTGTCGGT
GTTCAGTGCCGCCGTGGTGGTTGATCACGACGACCAGGACGAATCGCTGTTGGGGCAT
GGCGACCTGCGCCGCATCCCGACCCTGTATCCGGGCGAGCAGCAACTACCGACCGGC
CCCGGCGAGGAGCCGCCCAGCCAGCCCGGCATTCCGGGCATGGAACCAGACCTGCCA
GCCTTGACCGAAACGGAGGAATGGGAACGGCGCGGGCAGCAGCGCCTGCCGATGCCC
GATGAGCCGTGTTTTCTGGACGATGGCGAGCCGTTGGAGCCGCCGACACGGGTCACG
CTGCCGCGCCGGTAGCACTTGGGTTGCGCAGCAACCCGTAAGTGCGCTGTTCCAGACT
ATCGGCTGTAGCCGCCTCGCCGCCCTATACCTTGTCTGGGTA

Part 4

Antibiotic resistant cassette:

PmeI_CmR: pBBR1MCS-2

GGTAGTTTAAACTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTTTCGA
ATTTCTGCCATTCATCCGCTTATTATACTTATTCAGGCGTAGCACCAGGCGTTTAAGGGC
ACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTA
ATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAGACGGCATGATGAACCTGAATC
GCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGG
GGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGG
GATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTT
CACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGG
TATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGG
GTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGAT
GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT
TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT
TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG
GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATA
ACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTA
CGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA
GGGACACCGAAC

Replication sequence:

PmeI_pMB1: pBR322 (*Rop_ori*)

GGTAGTTTAAACTCCAGTAACCGGGCATGTTCATCATCAGTAACCCGTATCGTGAGCAT
CCTCTCTCGTTTCATCGGTATCATTACCCCCATGAACAGAAATCCCCCTTACACGGAGG
CATCAGTGACCAAACAGGAAAAAACCGCCCTTAACATGGCCCGCTTTATCAGAAGCCAG
ACATTAACGCTTCTGGAGAAACTCAACGAGCTGGACGCGGATGAACAGGCAGACATCT
GTGAATCGCTTCACGACCACGCTGATGAGCTTTACCGCAGCTGCCTCGCGCGTTTCGG
TGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTG
TAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGG
GTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTA
ACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCG

CACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGA
CTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTA
ATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCC
AGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCG
CCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACA
GGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCC
GACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTT
TCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGG
CTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGT
CTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACA
GGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAA
CTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCT
TCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGG
TTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTT
GATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAGAAC

PmeI_ pMB1: pUC19 (*ori*)

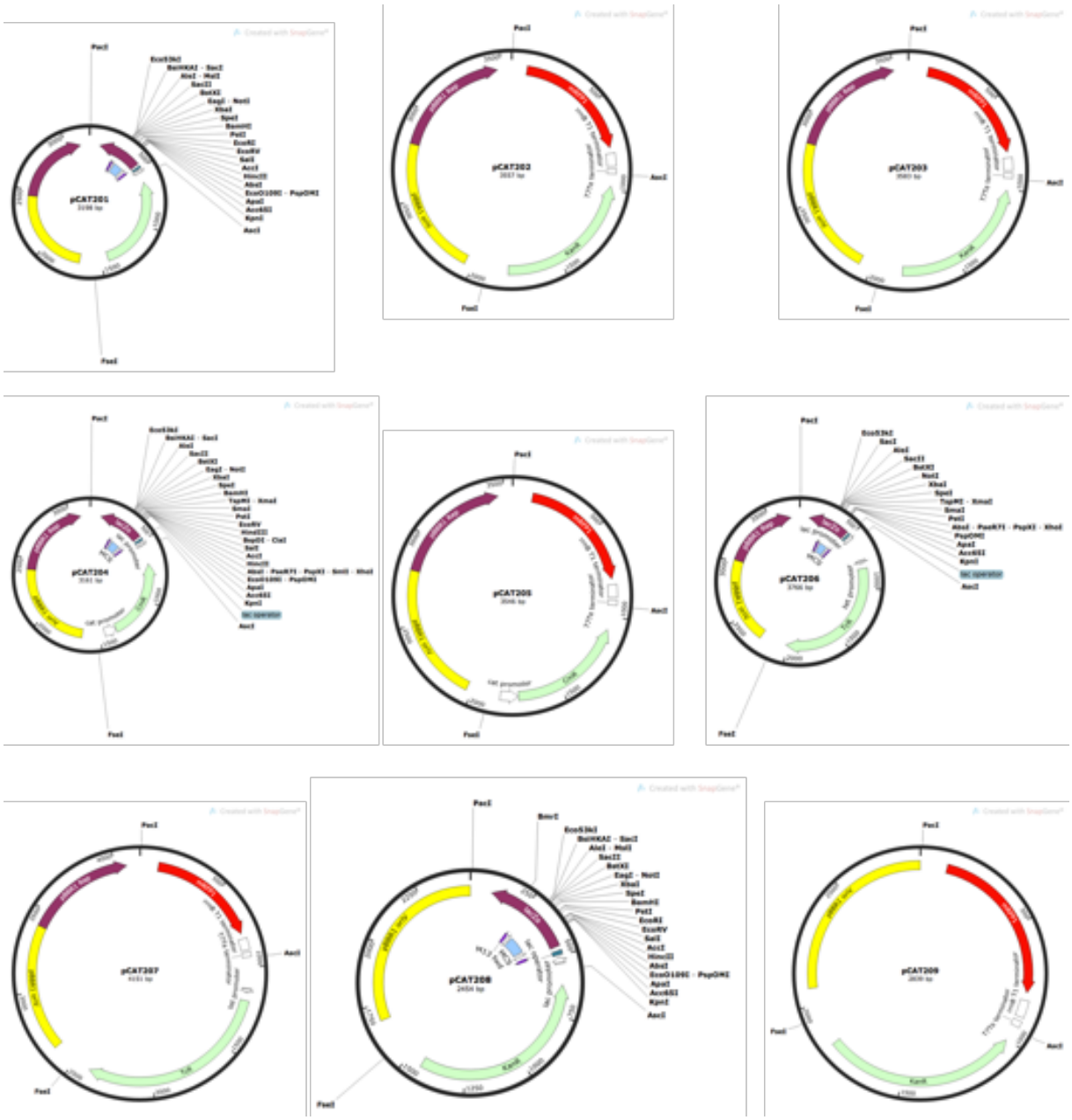
GGTAGTTTAAACGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCG
TTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTC
AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGA
AGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTT
TCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGG
TGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCG
CTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC
CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTAC
AGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCT
GCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAA
CAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAA
AAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCGAAC

PmeI_p15A: pBW115lac_hrpR

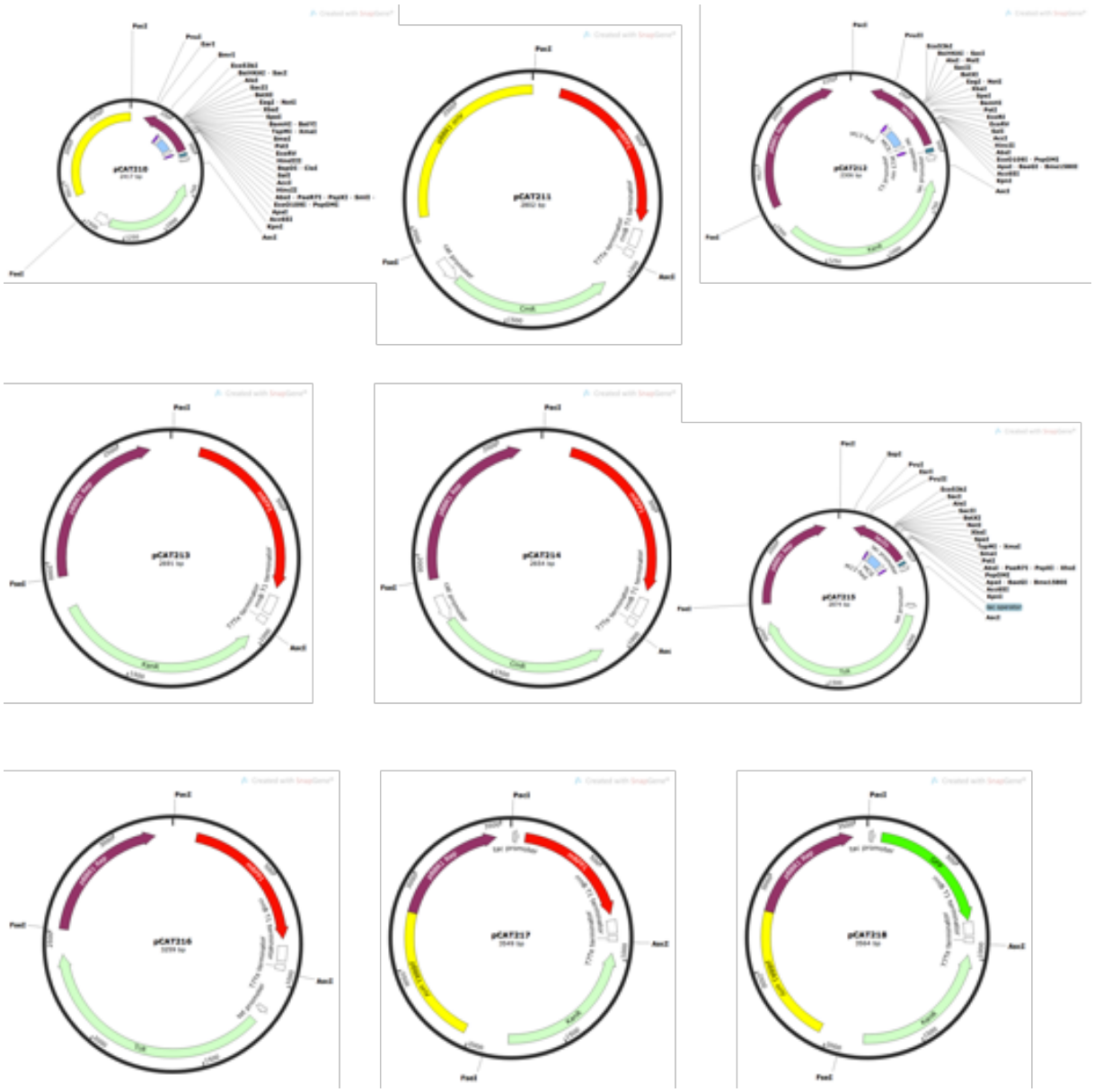
GGTAGTTTAAACCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACC
AAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAG
GTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACT
GAGCGTCAGACCCCTTAATAAGATGATCTTCTTGAGATCGTTTTGGTCTGCGCGTAATCT
CTTGCTCTGAAAACGAAAAAACCGCCTTGCAGGGCGGTTTTTCGAAGGTTCTCTGAGCT
ACCAACTCTTTGAACCGAGGTAACTGGCTTGGAGGAGCGCAGTCACCAAAACTTGTCCT
TTCAGTTTAGCCTTAACCGGCGCATGACTTCAAGACTAACTCCTCTAAATCAATTACCAG
TGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTCAAGACGATAGTTA
CCGGATAAGGCGCAGCGGTCGGACTGAACGGGGGGTTCGTGCATACAGTCCAGCTTG
GAGCGAACTGCCTACCCGGAACTGAGTGTCAGGCGTGGAATGAGACAAACGCGGCCAT
AACAGCGGAATGACACCGGTAAACCGAAAGGCAGGAACAGGAGAGCGCACGAGGGAG
CCGCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCACTGATT
TGAGCGTCAGATTTCGTGATGCTTGTCAGGGGGGCGGAGCCTATGGAAAAACGGCTTT
GCCGCGGCCCTCTCACTTCCCTGTTAAGTATCTTCCTGGCATCTTCCAGGAAATCTCCG
CCCCGTTCGTAAGCCATTTCCGCGAAC

Map of pCAT Plasmids

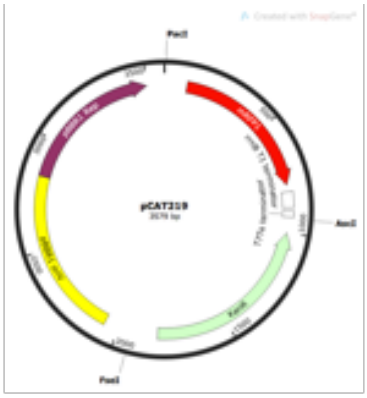
pCAT201–209

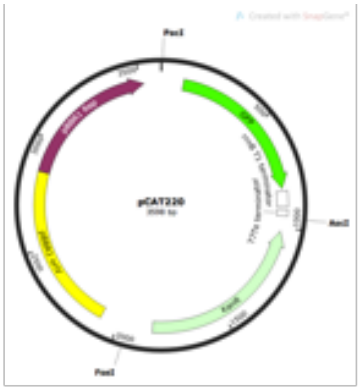


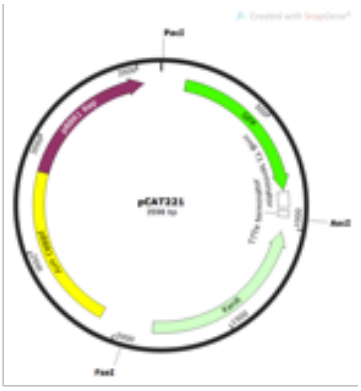
pCAT210–218

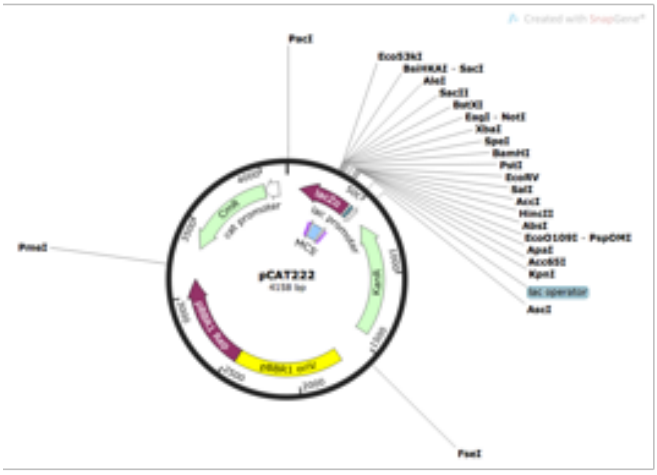


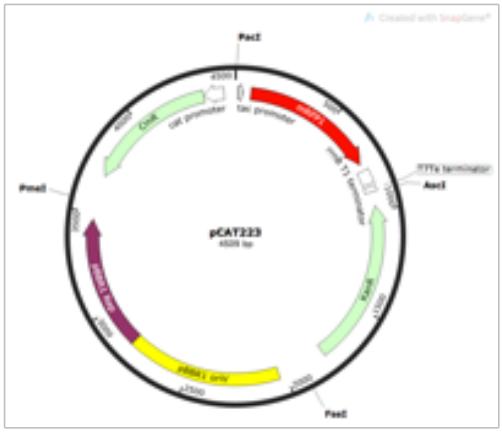
pCAT218–225

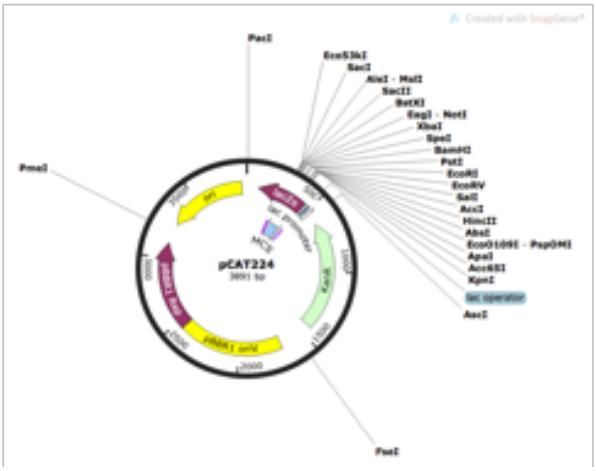

PacI
pCAT219
AscI
PmeI


PacI
pCAT220
AscI
PmeI


PacI
pCAT221
AscI
PmeI


PacI
pCAT222
PmeI
AscI
PmeI


PacI
pCAT223
PmeI
AscI
PmeI


PacI
pCAT224
PmeI
AscI
PmeI


PacI
pCAT225
PmeI
AscI
PmeI

Appendix D

RNA polymerase sigma factor RpoD

Ralstonia eutropha H16 chromosome 1

GenBank: AM260479.1

```
LOCUS       AM260479             4052032 bp    DNA     circular BCT 07-MAR-2015
DEFINITION  Ralstonia eutropha H16 chromosome 1.
ACCESSION   AM260479 REGION: complement(1..4052032)
VERSION     AM260479.1
DBLINK      BioProject: PRJNA13603
            BioSample: SAMEA3283071
KEYWORDS    complete genome.
SOURCE      Cupriavidus necator H16
  ORGANISM  Cupriavidus necator H16
            Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
            Burkholderiaceae; Cupriavidus.
REFERENCE   1
  AUTHORS   Pohlmann,A., Fricke,W.F., Reinecke,F., Kusian,B., Liesegang,H.,
            Cramm,R., Eitinger,T., Ewering,C., Potter,M., Schwartz,E.,
            Strittmatter,A., Voss,I., Gottschalk,G., Steinbuchel,A.,
            Friedrich,B. and Bowien,B.
  TITLE     Genome sequence of the bioplastic-producing 'Knallgas' bacterium
            Ralstonia eutropha H16
```

Two rpoD genes present in *C. necator* H16 (DNA binding region highlighted yellow)

Q0KB63 (656 AA) https://www.uniprot.org/uniprot/Q0KB63

MTSGVQRTAAKTRETGAIKDPGQTSAGSTAQPDPSTEAAARSQQLRALIQLGRQRGYLTHADISDHLPENFTDTAAMESIVSTFAEMGVKIYEQ
TPDAETLLLSDGPVVASDDQADEEAEVALATVDSEFGRTTDPVRMYMREMSSATLLTRKQEVEIAKRIEEGLNNMVHAISACPFTIAAILELSG
KVASNEISIDDLVDGLSDESIAEAAVAAAADETDDSVDSADEADDDSEDSDDEGSTQQSNEKALAQLREECLKRFARVTAQFELMCQESAANGA
GSAAFLAARDAVREELRTIRFTAKTIERLCANVQAMVDEVRTVERQVVQLLVERCGMEREEVIARFPGNETNLAWGQELVAQSRPYSAAVARAL
PDLEAHQQKLIDIQARAALSLPDLKGVNRKMLAAERQMRQAKHEMTQANLRLVISIAKKYTNRGMLFLDLIQEGNIGLMKAVDKFEYRRGWKFS
TYATWWVRQAVTRAIADQARTIRVPVHMIEQINKLNRLSREIMQQTGKEPDPAVLAERLDMTEDKVRSIMKIAKEPVSMETPVGEDGDTSLGDM
IADSDTATPADAALQAGLRAVVREMLDELTPREAKVLRMRFGIDMSTDYTLEEVGKQFDVTRERIRQIESKAMKKLRHPSRADQLITYLRDA

Q0K867 (827 AA) https://www.uniprot.org/uniprot/Q0K867

MKTTAAKTSTASVKTPARSAPSKSVRSAQAPAVTTPARTRTSEKTKASTSARESGKSASTSGKVAASAPVKGKSITVAKQQNTEVESKRAAAAG
AKSGEAKAGTTTTATKARAATPASVASERPAAPAIKPEPKKRGRKPKAEMQHDDSTTDDVTEEFYENDARPAATPAAAPKTEKQKAKDRKAKEK
ALLKEFASTQQGTEEELELRRQKLKALIKLGKSRGYLTYAEINDHLPDDMVDSETIDTLVATLNDIGIAVYEQAPDAETLLLNDNAPSATSEEE
AEEEAEAALSTVDSEFGRTTDPVRMYMREMGTVELLTREGEIEIAKRIEAGLKDMVMAISACPVTISEILAHAERVANDEIKIDEFVDGLIDPN
ADEAPEAPAAPAAAADDEDIESDDEEEGDEDDDDEGGAGAGASARQLEELKQNALEKFRVIAEQFDKMRRAFEKEGYNSKPYVKAQEAIQAELM
GIRFTARNVERLCDTLRGQVDEVRKLERSILNIVVDKCGMPRSEFVARFPGNETNLEWVHTIVADGKGYSTIVERNVPAVHELQQKLIDLQSRV
VLPLKELKGVNRKMAEGERRAREAKREMTEANLRLVISIAKKYTNRGLQFLDLIQEGNIGLMKAVDKFEYRRGYKFSTYATWWIRQAITRSIAD
QARTIRIPVHMIETINKMNRISRQILQETGNEPDPATLAEKMEMPEDKIRKIMKIAKEPISMETPIGDDDDSHLGDFIEDTNTLAPAEAALHGS
MRDVVKDVLDSLTPREAKVLRMRFGIEMSTDHTLEEVGKQFDVTRERIRQIEAKALRKLRHPSRSDKLKSFLEGN

Q0K867: The codon for the start amino acid is GTG (valine, V) instead of ATG (methionine, M).

Fasta file of desired amplicon for Q0KB63 (2352bp); starts immediately adjacent to the start codon (albeit on the complementary strand) of the upstream gene and end immediately prior to the start codon of the downstream gene.

>Q0KB63

TAAGCAGGTCTCGGAACTTAATTAAgatgctcctccttgaccggtgcggcccggtggcttccggcccacacttccagttgtaggacgacaa
cgcgcgaacgcaaagcacgcgcagctgtggcgcaccggctaccctcaactgtacgccggctccaccgggatacggcgagttgacgctgcgaagc
agcaatggctgtagaatacgcgtcctccccgattccaatggtcacggcgtccgatatcggcccgcgccgcaccggtacctcacgagccttgaca
gcaggcccgatagcgcacgcacagcgtgcgcgctgctgtcgaaacgcgattcacgaaggaaacagtagtaaATGacaagcggcgtacaacgaac
ggcagctaaaacacgagaaacaggcgcaatcaaggaccctggccagaccagcgcagggtcgacggcgcagcctgatccatcgacggaagccgcc
gcgcgtagccagcaattgcgtgcgctgatccagctgggcaggcagcgcggctacctgacccacgcggatatcagcgaccacctgccggagaact
tcactgacacggcggcgatggaaagcatcgtcagcacgttcgccgagatgggcgtgaagatctatgagcagacgccggatgccgagacgctgct
cctgagcgatggcccggtggtggcctcggacgaccaggccgatgaagaggccgaggtagcgcttgcgactgtggactccgagttcggccggacc
accgacccggtgcgcatgtacatgcgcgaaatgagctcggcaacgctgctcacgcgcaagcaggaagtcgagatcgccaagcgcatcgaggaag
gcctgaacaatatggtgcatgccatctcggcctgcccgttcaccatcgccgcgattctcgagctgtcgggcaaggtcgccagcaatgaaatcag
catcgacgacctggtcgatggcctgagcgacgagagtattgccgaagcagccgtggcggcggcagctgacgagaccgacgacagcgtcgatagt
gccgacgaggccgatgacgattccgaagactccgacgacgaaggctcgacccagcaaagcaacgaaaaagccctggcccaactgcgtgaggaat
gcctgaagcggttcgcgcgcgtcaccgcgcaattcgagctgatgtgccaggaaagcgccgcaaatggcgcgggctcggcagccttcctggccgc
aagggatgcggtgcgcgaagaactgcgcacgatccgctttaccgcgaagaccatcgagcgcctgtgcgccaacgtccaggccatggtcgacgag
gtgcgcacggtcgagcgccaggtggtccagctgctggtcgagcgctgcgggatggagcgcgaggaggtcatcgcccgcttccccggcaatgaaa
ccaacctggcctgggggcaggagctggtcgcccagtcccgtccgtacagcgcggcggtcgcgcgggccctgccggacctcgaggcgcaccagca
aaagctgatcgacatccaggcccgggcggcgctgtcgctgcccgacctgaagggcgtcaaccgcaagatgctggctgccgagcggcagatgcgc
caggccaagcacgagatgacgcaggccaacctgcggctggtgatctcgattgccaagaagtacaccaaccgcgggatgctgttcctggatctga
tccaggaaggcaacatcggcctgatgaaggcggtcgacaagttcgaatatcgccgcggctggaaattctccacctacgcgacgtggtgggtgcg
ccaggccgtgacgcgggcgattgccgaccaggcccgcaccatccgcgtcccggtgcacatgatcgagcagatcaacaagctcaatcgcctgtcg
cgcgagatcatgcagcagaccggcaaggaacccgatccggcggtgctggccgagcgcctggacatgaccgaagacaaggttcgctcgatcatga
aaatcgcgaaagagccggtctcgatggaaaccccggtcggcgaggatggcgataccagcctgggcgacatgattgccgactcggacacggccac
gcccgccgacgccgcattgcaggcgggcctgcgcgccgtcgtgcgcgaaatgctcgacgagctcacgccgcgtgaagccaaggtgctgcggatg
cgctttggcatcgatatgtccaccgactacacgctggaagaagtcggcaagcagttcgatgtcacgcgcgagcggatccgccagattgaaagca
aggcaatgaagaagctgaggcatcccagccgggcagaccagctgatcacctatctgcgcgacgcctgatacttgccggcaacagccgcttggcc
gactgacacagagcagacaggagctttcggAAGCGGAGACCTAAGCA

Forward primer: 5’-TAAGCAGGTCTCGGAACTTAATTAAGATGCTCCTCCTTGACCGGT-3’
Reversed primer: 5’-TGCTTAGGTCTCCGCTTCCGAAAGCTCCTGTCTGCTC-3’

Fasta file of desired amplicon for Q0K867 (2905bp); starts immediately after the stop codon of the upstream gene dnaG and ends immediately prior to the stop codon of the downstream gene plcN1 (on the complementary strand).

>Q0K867

TAAGCAGGTCTCGGAACTTAATTAA**cttggctga**ggtgcctgtgtgcgccctggtcatcaaccaggaggtggcagtaagcgttggataagg
gactgggcccggccccgccgagcttaagccttgatttttcaggtgaaaacctcacctgatagggagccggctatgcgtggcgtgctacaatagc
cggtttggattgcgaaatctttctctccagttctggtgaaagtgagcgtgccaatggcaaaggccaaagcaaccgaaagggttacctctaaagc
tccccaatc**agggagcggcaag**gcgtct**GTG**aaaacaacggcggcgaagacatcgaccgcgtcagtcaagacaccagcccggtctgcgccatca
aagtccgtgcgtagcgcgcaagcgcccgccgtgacgactcccgccaggactcgcacatccgaaaagacaaaggcctcgacttcggcacgcgaga
gcggcaagagcgccagtacgagcggcaaggtcgccgcatctgcaccagtgaaaggcaagagtatcaccgtggcgaaacagcagaataccgaagt
cgagagcaagcgagcggcagccgccggcgccaagagtggggaagccaaggctggcaccaccactacggcaaccaaggcgcgtgccgcgacacct
gcatccgttgcatccgagcgtcccgctgcgccggcaatcaagccggagccgaaaaagcgtggccgcaagcccaaggccgagatgcagcacgacg
acagcacaacagacgacgtgactgaagagttttacgagaacgatgcgcgcccggcggcaaccccggccgcggcgccgaagaccgaaaagcagaa
agccaaggaccgcaaggccaaggaaaaggccctgctcaaggagttcgcctcgacccagcagggtaccgaggaagagcttgagctgcgtcgccag
aaactcaaggcgctgatcaagctgggcaagtcgcgcggctacctgacctacgcggaaatcaacgatcacctgccggacgacatggtcgattcgg
aaacgatcgacacgctggtcgccacgctgaacgacatcggcatcgctgtctacgaacaggcgccggatgccgagacgctgctgctcaacgacaa
cgccccgtccgccaccagcgaggaagaagccgaggaagaggccgaagcggccctgtccacggtggactccgagttcggccgcaccaccgacccg
gtgcgcatgtacatgcgcgaaatgggcacggtcgagctgctcacgcgcgaaggcgaaatcgagatcgccaagcgcatcgaggccggcctgaagg
acatggtgatggcgatttcggcttgcccggtcaccatctccgagatcctggcccatgccgagcgcgtggccaacgacgagatcaagatcgacga
attcgtcgacggcctgatcgacccgaacgccgacgaagccccggaagcgcccgcagccccggctgcagcggccgacgacgaggatatcgagtcc
gacgacgaagaggaaggcgacgaggacgatgacgacgaaggcggcgccggcgccggtgcctccgcgcgccagctggaagaactgaagcagaacg
cgctggagaaattccgcgtgatcgccgagcagttcgacaagatgcgccgcgcgttcgaaaaggaaggctacaactccaagccctacgtcaaggc
gcaggaagccatccaggccgagctgatgggcatccgcttcaccgcccgcaatgtcgagcgcctgtgcgacacgctgcgcggccaggtcgacgaa
gtgcgcaagctcgaacgctcgatcctgaacatcgtggtcgacaagtgcggcatgccgcgctcggaattcgtcgcccgcttcccgggcaacgaga
ccaacctcgagtgggtccacaccatcgtcgccgacggcaagggctacagcaccatcgtcgagcgcaacgtgccggccgtgcatgagctgcagca
gaagctgatcgacctgcagtcgcgcgtggtgctgccgctgaaggaactgaagggcgtcaaccgcaagatggccgagggcgagcgtcgtgcccgc
gaagccaagcgcgagatgaccgaggccaacctgcgtctggtgatttcgatcgccaagaagtacacgaatcgcggcctgcagttcctcgacctga
tccaggaaggcaacatcggcctgatgaaggcggtggacaagttcgaataccgccgcggctacaagttctcgacctacgccacgtggtggatccg
ccaggccatcacgcgctcgatcgccgaccaggcgcgcaccatccgtatcccggtgcacatgatcgagaccatcaacaagatgaaccgcatctcg
cgccagatcctgcaggaaaccggcaacgagccggatccggcaacgctggccgagaagatggagatgccggaagacaagatccgcaagatcatga
agatcgccaaagagccgatctccatggaaacgccgatcggtgacgacgacgactcccatctgggcgacttcatcgaggacaccaacacgctggc
cccggccgaagccgcgctgcacggctccatgcgcgacgtcgtcaaggacgtgctggactcgctcacgccgcgcgaagccaaggtgctgcgcatg
cgcttcggcatcgaaatgagcaccgaccacacgctggaagaggtcggcaagcagttcgacgtcacgcgtgaacggatccgccagatcgaggcca
aggcactgcgcaagctgcgccacccgagccgctcggacaagctgaagagcttcctggaaggcaac**taa**acgcttccaagggcctctagctcatg
cctggttagagcagcggactcataatccgttggtgccgtgttcgactcacgggaggcccaccaatgtgaaaaaggactgcaggttgatacctgc
agtcctttttttt**gcctcgaactaccgctcaat**AAGCGGAGACCTAAGCA

Forward primer: 5’-TAAGCAGGTCTCGGAACTTAATTAA**CTTGGCTGA**GGTGCCTGTG-3’
Reverse primer: 5’-TGCTTAGGTCTCGCTTATTGAGCGGTAGTTCGAGGC-3’

N.B. Genomic DNA sequence displayed in reverse complement

326 gap between predicated glutamine amidotransferase and rpoD1; 57 gap between rpoD and hypothetical membrane associated protein

Genetic context for Q0KB63

```
     gene            complement(2280405..2281250)
                     /gene="h16_A1627"
                     /locus_tag="H16_A1627"
     CDS             complement(2280405..2281250)
                     /gene="h16_A1627"
                     /locus_tag="H16_A1627"
                     /EC_number="2.-.-.-"
                     /function="General function prediction only"
                     /inference="protein motif:CDD:cd01908"
                     /codon_start=1
                     /transl_table=11
                     /product="predicted glutamine amidotransferase"
                     /protein_id="CAJ92759.1"
                     /db_xref="EnsemblGenomes-Gn:H16_A1627"
                     /db_xref="EnsemblGenomes-Tr:CAJ92759"
                     /db_xref="GOA:Q0KB62"
                     /db_xref="InterPro:IPR017932"
                     /db_xref="InterPro:IPR026869"
                     /db_xref="InterPro:IPR029055"
                     /db_xref="UniProtKB/TrEMBL:Q0KB62"
                     /translation="MCRWLAYSGKSVPLETVLFQPEHSLIDQSLNSRLGHTTTNGDGF
                     GVGWYGRHSEVPFRYRCLHPAWSDTNLRETARAVRAPLFVAHVRAATGTPTQETNCHP
                     FRFGRWLFVHNGLIREYPLLRRDLMLAVAPHLFRWIEGSTDSEVMFFLALSFGLERDP
                     GGALEQMAGLIEDVGHRYDVQFPLNMTVCVTDGQQIVAVRYSSETDSRSLFHSTSFRQ
                     LRALYPDDPRIIAAGDNAFLVLSEPLIEVQGVWEEIPEATTLVARGGEIERRRFVPRM
                     PSQSA"

     gene            2281576..2283546
                     /gene="rpoD1"
                     /locus_tag="H16_A1626"
     CDS             2281576..2283546
                     /gene="rpoD1"
                     /locus_tag="H16_A1626"
                     /EC_number="2.7.7.6"
                     /function="Transcription"
                     /inference="protein motif:COGS:COG0568"
                     /inference="protein motif:COGS:COG1191"
                     /inference="protein motif:PFAM:PF03979"
                     /inference="protein motif:PFAM:PF04539"
                     /inference="protein motif:PFAM:PF04542"
                     /inference="protein motif:PFAM:PF04545"
                     /inference="protein motif:PFAM:PF04546"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA-directed RNA polymerase sigma subunit
                     (RpoD)"
                     /protein_id="CAJ92758.1"
                     /db_xref="EnsemblGenomes-Gn:H16_A1626"
                     /db_xref="EnsemblGenomes-Tr:CAJ92758"
                     /db_xref="GOA:Q0KB63"
                     /db_xref="InterPro:IPR000943"
                     /db_xref="InterPro:IPR007127"
                     /db_xref="InterPro:IPR007624"
                     /db_xref="InterPro:IPR007627"
                     /db_xref="InterPro:IPR007630"
                     /db_xref="InterPro:IPR007631"
                     /db_xref="InterPro:IPR009042"
                     /db_xref="InterPro:IPR011991"
                     /db_xref="InterPro:IPR012760"
                     /db_xref="InterPro:IPR013324"
                     /db_xref="InterPro:IPR013325"
                     /db_xref="InterPro:IPR014284"
```

```
                /db_xref="InterPro:IPR028630"
                /db_xref="UniProtKB/TrEMBL:Q0KB63"
                /translation="MTSGVQRTAAKTRETGAIKDPGQTSAGSTAQPDPSTEAAARSQQ
                LRALIQLGRQRGYLTHADISDHLPENFTDTAAMESIVSTFAEMGVKIYEQTPDAETLL
                LSDGPVVASDDQADEEAEVALATVDSEFGRTTDPVRMYMREMSSATLLTRKQEVEIAK
                RIEEGLNNMVHAISACPFTIAAILELSGKVASNEISIDDLVDGLSDESIAEAAVAAAA
                DETDDSVDSADEADDDSEDSDDEGSTQQSNEKALAQLREECLKRFARVTAQFELMCQE
                SAANGAGSAAFLAARDAVREELRTIRFTAKTIERLCANVQAMVDEVRTVERQVVQLLV
                ERCGMEREEVIARFPGNETNLAWGQELVAQSRPYSAAVARALPDLEAHQQKLIDIQAR
                AALSLPDLKGVNRKMLAAERQMRQAKHEMTQANLRLVISIAKKYTNRGMLFLDLIQEG
                NIGLMKAVDKFEYRRGWKFSTYATWWVRQAVTRAIADQARTIRVPVHMIEQINKLNRL
                SREIMQQTGKEPDPAVLAERLDMTEDKVRSIMKIAKEPVSMETPVGEDGDTSLGDMIA
                DSDTATPADAALQAGLRAVVREMLDELTPREAKVLRMRFGIDMSTDYTLEEVGKQFDV
                TRERIRQIESKAMKKLRHPSRADQLITYLRDA"

gene            2283603..2284187
                /gene="h16_A1625"
                /locus_tag="H16_A1625"
CDS             2283603..2284187
                /gene="h16_A1625"
                /locus_tag="H16_A1625"
                /codon_start=1
                /transl_table=11
                /product="hypothetical membrane associated protein"
                /protein_id="CAJ92757.1"
                /db_xref="EnsemblGenomes-Gn:H16_A1625"
                /db_xref="EnsemblGenomes-Tr:CAJ92757"
                /db_xref="InterPro:IPR021212"
                /db_xref="UniProtKB/TrEMBL:Q0KB64"
                /translation="MTDSNLSFFGRVSLAMGTFFAILGNRELAAGVKRLRDGEGFAAA
                PAPAAVSAPAPAPVAEAAAPVLKEASPVAALQLLGLLQRDARFIDFVEEDIARYSDTE
                IGAAARLVHDGCRGVLREHFTIRPVREEAEGSRVTLADGFDATAIRLTGNVVGSAPFH
                GSISHRGWKVEEVRLPRVAERHDATVIAPAEVEL"
```

N.B. Genomic DNA sequence displayed in reverse complement

Genetic context for Q0K867
274 gaps between dnaG and rpoD1; 149 gap between rpoD1 and Phospholipase C

```
     gene            1106048..1107859
                     /gene="dnaG"
                     /locus_tag="H16_A2726"
     CDS             1106048..1107859
                     /gene="dnaG"
                     /locus_tag="H16_A2726"
                     /EC_number="2.7.7.-"
                     /function="DNA replication, recombination, and repair"
                     /inference="protein motif:COGS:COG0358"
                     /inference="protein motif:PFAM:PF01807"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA primase"
                     /protein_id="CAJ93805.1"
                     /db_xref="EnsemblGenomes-Gn:H16_A2726"
                     /db_xref="EnsemblGenomes-Tr:CAJ93805"
                     /db_xref="GOA:Q0K866"
                     /db_xref="InterPro:IPR002694"
                     /db_xref="InterPro:IPR006171"
                     /db_xref="InterPro:IPR006295"
                     /db_xref="InterPro:IPR013173"
                     /db_xref="InterPro:IPR013264"
                     /db_xref="InterPro:IPR016136"
                     /db_xref="InterPro:IPR019475"
                     /db_xref="UniProtKB/TrEMBL:Q0K866"
                     /translation="MIPQSFIQDLLNRVDIVDVVGKYVQLKKGGANFMGLCPFHNEKS
                     PSFTVSPTKQFYHCFGCGAHGSAIGFLMEFSGQSYPEAVRELAQSVGMSVPEERDRLP
                     PGQRAEQQARSVALSDAMTRATDFYRRQLRSAPQAIQYLKGRGLTGEIAANFGLGYAP
                     DDWQGLEAVFGSYRDDNVSAPLVECGLLIESDKRDADGKPRRYDRFRDRIMFPIRNTK
                     GAVIGFGGRVMGQGEPKYLNSPETPLFSKGTELYGLFEARHAIRETGYVLVVEGYMDV
                     VALAQLGFANAVATLGTACTPVHVQKLLRQTDAVVFSFDGDSAGRRAARRALEACLPH
                     VADNKTIRFLFLPAEHDPDSYVREEGTEAFATQVRNAMPLSRFLLQSVTEDLDLRQPE
                     GRARAQYEAKPLLQAMPAGGLRLQIVRELADATGTTPAEIEAICGLRSDPARVGRFAQ
                     PRPRTRRQAPTGLEQRVIQLLMCYPALSARLDEDARALLLGGESSDSEVLAHLVEACD
                     SVQGEVNFAAFSEHLAQSPYAEVYAVARAAVLREEIEEAPAIQDFDAAITKLLAEPLR
                     RELDGLQAEVVAGTADEAAKQRMRWLVGEIHRRRQLG"

     gene            1108133..1110616
                     /gene="rpoD1"
                     /locus_tag="H16_A2725"
     CDS             1108133..1110616
                     /gene="rpoD1"
                     /locus_tag="H16_A2725"
                     /EC_number="2.7.7.6"
                     /function="Transcription"
                     /inference="protein motif:COGS:COG0568"
                     /inference="protein motif:PFAM:PF03979"
                     /inference="protein motif:PFAM:PF04539"
                     /inference="protein motif:PFAM:PF04542"
                     /inference="protein motif:PFAM:PF04545"
                     /inference="protein motif:PFAM:PF04546"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA-directed RNA polymerase sigma subunit
                     (RpoD)"
                     /protein_id="CAJ93804.1"
                     /db_xref="EnsemblGenomes-Gn:H16_A2725"
                     /db_xref="EnsemblGenomes-Tr:CAJ93804"
                     /db_xref="GOA:Q0K867"
                     /db_xref="InterPro:IPR000943"
```

```
                     /db_xref="InterPro:IPR007127"
                     /db_xref="InterPro:IPR007624"
                     /db_xref="InterPro:IPR007627"
                     /db_xref="InterPro:IPR007630"
                     /db_xref="InterPro:IPR007631"
                     /db_xref="InterPro:IPR009042"
                     /db_xref="InterPro:IPR011991"
                     /db_xref="InterPro:IPR012760"
                     /db_xref="InterPro:IPR013324"
                     /db_xref="InterPro:IPR013325"
                     /db_xref="InterPro:IPR014284"
                     /db_xref="InterPro:IPR028630"
                     /db_xref="UniProtKB/TrEMBL:Q0K867"
                     /translation="MKTTAAKTSTASVKTPARSAPSKSVRSAQAPAVTTPARTRTSEK
                     TKASTSARESGKSASTSGKVAASAPVKGKSITVAKQQNTEVESKRAAAAGAKSGEAKA
                     GTTTTATKARAATPASVASERPAAPAIKPEPKKRGRKPKAEMQHDDSTTDDVTEEFYE
                     NDARPAATPAAAPKTEKQKAKDRKAKEKALLKEFASTQQGTEEELELRRQKLKALIKL
                     GKSRGYLTYAEINDHLPDDMVDSETIDTLVATLNDIGIAVYEQAPDAETLLLNDNAPS
                     ATSEEEAEEEAEAALSTVDSEFGRTTDPVRMYMREMGTVELLTREGEIEIAKRIEAGL
                     KDMVMAISACPVTISEILAHAERVANDEIKIDEFVDGLIDPNADEAPEAPAAPAAAAD
                     DEDIESDDEEEGDEDDDDEGGAGAGASARQLEELKQNALEKFRVIAEQFDKMRRAFEK
                     EGYNSKPYVKAQEAIQAELMGIRFTARNVERLCDTLRGQVDEVRKLERSILNIVVDKC
                     GMPRSEFVARFPGNETNLEWVHTIVADGKGYSTIVERNVPAVHELQQKLIDLQSRVVL
                     PLKELKGVNRKMAEGERRAREAKREMTEANLRLVISIAKKYTNRGLQFLDLIQEGNIG
                     LMKAVDKFEYRRGYKFSTYATWWIRQAITRSIADQARTIRIPVHMIETINKMNRISRQ
                     ILQETGNEPDPATLAEKMEMPEDKIRKIMKIAKEPISMETPIGDDDDSHLGDFIEDTN
                     TLAPAEAALHGSMRDVVKDVLDSLTPREAKVLRMRFGIEMSTDHTLEEVGKQFDVTRE
                     RIRQIEAKALRKLRHPSRSDKLKSFLEGN"

gene                 complement(1110765..1113011)
                     /gene="plcN1"
                     /locus_tag="H16_A2724"
CDS                  complement(1110765..1113011)
                     /gene="plcN1"
                     /locus_tag="H16_A2724"
                     /EC_number="3.1.4.3"
                     /inference="protein motif:PFAM:PF04185"
                     /codon_start=1
                     /transl_table=11
                     /product="Phospholipase C"
                     /protein_id="CAJ93803.1"
                     /db_xref="EnsemblGenomes-Gn:H16_A2724"
                     /db_xref="EnsemblGenomes-Tr:CAJ93803"
                     /db_xref="GOA:Q0K868"
                     /db_xref="InterPro:IPR006311"
                     /db_xref="InterPro:IPR007312"
                     /db_xref="InterPro:IPR008475"
                     /db_xref="InterPro:IPR017767"
                     /db_xref="UniProtKB/TrEMBL:Q0K868"
                     /translation="MVSRRNFLQAAAGTGFAAAALAAFPPSIRKALAIPANNATGTIK
                     DVEHVVILMQENRSFDHYFGTLKGVRGFGDRFTIPLPNARKVWQQQRSNGAVLTPYHL
                     DGTNNNAQRAAGTPHAWVDSQQAWDHGRMASWPTYKTNTSMGYFKEKEIPFQFALANA
                     FTLCDAYHCSMHTGTDANRSFHLTGTNGPTAANVAFVNNEWDAIDGLPASANTGYTWK
                     TYAERLEAAGISWICYQNMPDEWGDNMLGAFQQFRKANLASGFPVSSGGAPGAPYANT
                     GQPLPYHAYDAATDNAANPLYKGVANTLPGTRPEEYLDAFRRDIKEGRLPQVSWINAP
                     SIYCEHPGPSSPVQGAWFLQEVLDALTAVPEVWSKTVLLVNFDENDGYFDHVPSPSAP
                     SVNPDKTLAGKATLSDAEMQAEYFNHPPPPGSRTQPAADGRVYGPGPRVPLYAISPWS
                     RGGWINSQVFDHTSVLRFLEARFGVAEPNISPFRRAVCGDLTSAFNFKTPNSEALPTL
                     SGRTTRSGADQLRQAQQALPAVPLPVDMQLPLQATGTRPSRALPYELHTSARCSAVGQ
                     VELVFANTGTQAAVFHVYDRYQLGRIPRRYVVEARKSLSDTWNVFQDNAGQYDLWVLG
                     PNGFHRHFRGDTNRIGDTGIAPEIRVCYDIANGDVYVDLINAGRKACHFSIQALAYRT
                     DGPWPVTVGANDSKSVHWSLEESGQWYDFAVTCDSDPAFYRRFAGRVENGRHTVSDPA
                     MGMVTAQD"
```

Data Driven Modelling of a Chemically Defined Growth Medium for *Cupriavidus necator* H16

Christopher C. Azubuike, Martin G. Edwards, Angharad M. R. Gatehouse and Thomas P. Howard

School of Natural and Environmental Sciences, Faculty of Science, Agriculture and Engineering, Newcastle University, Newcastle-upon-Tyne, United Kingdom

Cupriavidus necator **is a Gram-negative soil bacterium of major biotechnological interest. It is a producer of the bioplastic 3-polyhydroxybutyrate, has been exploited in bioremediation processes, and it's lithoautotrophic capabilities suggest it may function as a microbial cell factory upgrading renewable resources to fuels and chemicals. It remains necessary however to develop appropriate experimental resources to permit controlled bioengineering and system optimisation of this microbial chassis. A key resource for physiological, biochemical and metabolic studies of any microorganism is a chemically defined growth medium. Here we use 1 mL micro-well cell cultures, automated liquid handling and a statistical engineering approach to develop a model that describes the effect of key media components and their interactions on *C. necator* culture cell density. The model is predictive and was experimentally validated against novel media compositions. Moreover, the model was further validated against larger culture volumes at 100 mL and 1 L volumes and found to correlate well. This approach provides valuable and quantifiable insights into the impact of media components on cell growth as well as providing predictions to guide culture scale-up.**

Cupriavidus necator **H16 | Design of Experiments (DOE) | chemically defined media | 3-polyhydroxybutyrate (PHB)**

Correspondence: ***thomas.howard@newcastle.ac.uk***

Introduction

The Gram-negative, non-spore forming soil bacterium, *Cupriavidus necator* H16, is of biotechnological importance principally due to its ability to accumulate >80% of its dry cell weight as polyhydroxyalkanoate (PHA) - a biodegradable polymer and an alternative to petroleum-based polymers (1, 2). The accumulation of PHA (specifically 3-polyhydroxybutyrate (3-PHB)) by the bacterium is a carbon conservation mechanism. When nitrogen, oxygen, or phosphorus becomes limiting to cell growth the bacterium diverts excess carbon to 3-PHB. This carbon store supports growth when conditions improve (3–6). *C. necator* H16 is also chemolithoautotrophic, with the ability to use CO_2, and formate or H_2 as carbon and energy sources to support cellular metabolism (7). In the absence of oxygen it can use alternative electron acceptors (NO_3^- and NO_2^-) to carry out anaerobic respiration by denitrification (8, 9). Metabolic engineering of *C. necator* H16 has been demonstrated through the introduction of pathways for the biosynthesis of alcohols (6, 7, 10, 11), fatty acids (12–14), alkanes (15) and enzymes (16) under both heterotrophic and autotrophic growth conditions. The production of branched chain alcohols by *C. necator* H16 through electricity powered cellular synthesis (i.e. electrosynthesis), demonstrates the value of this bacterium as a chassis capable of exploiting renewable feedstock for the biosynthesis of valuable products (7). More widely, the genus is known to encode genes facilitating metabolism of environmental pollutants such as aromatics and heavy metals making it a potential microbial remediator (8, 9, 17–19).

As with any bacterium, the development of *C. necator* as an industrial chassis requires appropriate tools for studying and engineering the organism. One of these resources is the availability of a characterised, chemically defined growth medium. Chemically defined media are important to enable experimental reproducibility, to reliably characterise the genetics of the organism, to determine genotype by environment interactions, and to facilitate fundamental research of bacterial physiology that underpins bioengineering efforts. While different chemically defined media have been deployed for the cultivation of *C. necator* H16 (3, 10–12, 20–22) there is no consensus regarding the components that are required, the concentration of each component, or how each component interacts to affect growth of the bacterium (SI Table 1).

Design of Experiments (DOE) is an iterative, empirical approach that systematically explores the relationship between input variables (factors) and output variables (responses). The approach yields a structured set of data that can be used to build statistical models employed in understanding or optimising system performance. These statistical models can be validated against prior knowledge, internal statitical methods or ultimately - by their ability to predict responses from new combinations of factors. Despite its early origins within the biological sciences, DOE is not a common-place method for life scientists. Increasing availability of laboratory automation and high throughput technologies may be changing this. DOE has found use in the optimisation of metabolic pathways (23, 24), cell-free protein synthesis reactions (25) and codon-use algorithms (26). It has been applied to the study of genotype-by-genotype and genotype-by-environment interactions in yeast (27) and in re-purposing enzyme activities (28).

Here we employ a statistical engineering approach to build a data driven model that can accurately predict *C. necator* H16 growth responses to a range of media formulations. The model highlighted different formulations that allow reproducible and robust growth of *C. necator* H16 with the minimal concentration of each component and allowed us to identify and understand interactions between components of the media. Finally, the model allowed the learning from the small scale (1 mL) to be applied at larger volumes (100 mL and 1

L). Understanding the impact of each component of a chemically defined medium on *C. necator* growth is a fundamental tool for controlled exploration of the biotechnological potential of this important bacterium.

Results

Identifying main ingredients in chemically-defined media that affect the growth of *C. necator*. A comparison of the literature for the use of defined media for *C. necator* growth identified variety in both the nature and concentrations of macroelements (C, N, P, Ca, Mg, and S) and trace elements (Cu, Zn, Fe and Mn) required for robust cell growth (Table SI 1). An initial scoping trial was carried out using fructose, glucose, glycerol or sucrose to identify a principle carbon source to be used for subsequent work, and to determine the range of concentrations to be tested. Four scoping trials were performed one each at low and high concentrations of media components, and two trials at the midpoint values between the two extremes. Cultures were grown in 1 mL volumes in a 48-well plate format for 120 h, at 30°C, 200 rpm. We confirmed that optical density (OD_{600nm}) was an appropriate surrogate for determining *C. necator* cell growth (Fig. SI 1A). At the ranges tested glucose, glycerol and sucrose supported little or no growth of *C. necator* (Fig. SI 1B-D), whilst fructose supported high growth except at the lowest concentrations (Fig. 1). The OD_{600nm} for the two midpoint experiments demonstrated a peak at 72 h followed by a plateau. Growth obtained from optical density measurements at $OD_{600\ nm}$ correlated well with the number of viable cells (CFU/mL). From these scoping trails it was established that fructose would be our principle carbon source, OD_{600nm} was an appropriate measure of cell growth, growth assays in 1 mL volumes in a 48-well plate format were appropriate for subsequent experiments, and recording $OD_{600\ nm}$ at 72 h provides a good balance between measuring growth rate and peak culture density.

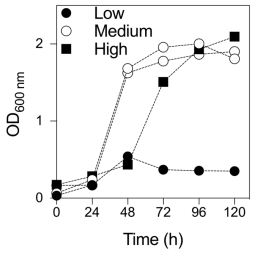


Fig. 1. Scoping trials. Growth of *C. necator* H16 in 48-well format with fructose as the carbon source. Each experiment was carried out at low, medium ($n = 2$) or high concentrations of each media component (details can be found in Table SI 2).

We next identified key factors that might influence OD_{600nm} at 72 h. A definitive screening design (DSD) array was built based on 10 media components (Table 1). DSDs are highly efficient experimental designs in which all main effects can be estimated independently of other main effects and all possible two-way interactions. The requisite variant media compositions were assembled in a 48-well plate using a liquid-handling robot. Cell growth in these variant media was monitored as above. The resulting data were analysed using both the *Definitive Screening* and the *Two Level Screening* platforms in JMP Pro 13.0. Both analyses indicated that fructose, $CaCl_2$ and amino acids contributed positively to growth, while Na_2HPO_4 and trace elements contribute negatively to growth. Factors such as NaH_2PO_4, K_2SO_4, $MgSO_4$, NH_4Cl and vitamins were not found to have statistically significant effects under the conditions tested (Fig. 2). These analyses also highlighted several two-way interactions that may influence growth responses, however, while DSDs are efficient arrays for de-aliasing main effects from other main effects and from second-order interactions, second-order interactions remain partially aliased with themselves. For this reason attributing effects to specific second-order interactions was not attempted at this stage.

Augmentation of the data set. A definitive screening design can force many of the data-points collected to the edges of the design space. For this reason we ran a second DSD array (Table SI 3) in which the concentration ranges of the components were guided by data from DSD1 and the factors under investigation were restricted to those that were highlighted as significant in DSD1. Examining the combined data for DSD1 and DSD2 did indeed confirm that the extreme concentrations of some of the components were detrimental to cell growth. For example, when the concentrations of fructose were at the highest and lowest values (0.5 and 40 g/L) the $OD_{OD600nm}$ 72 h were both lower and more variable than when fructose was restricted to between 5 and 25 g/L (Fig. 3). This indicates that maintaining fructose between 5 and 25 g/L is key to establishing robust and reliable growth. Likewise, adjustments were made to the concentration ranges of the amino acids (5 and 20 ml/L), $CaCl_2$ (0.1 and 0.459 g/L) and Na_2HPO_4 (0.1 and 3.05 g/L). Other factors not identified as statistically significant were kept at the midpoint from DSD1, with the exception of NH_4Cl which was set at the

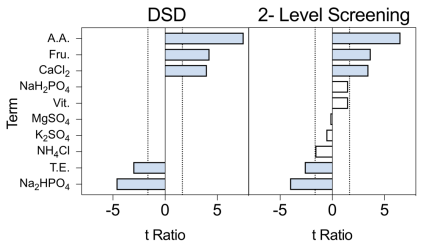


Fig. 2. Definitive Screening Design array analysis. Definitive screening and 2-Level screening of data within DSD1 were performed. The comparative lengths of the t-ratios for each factor and factor interaction are shown. Bars extending to the right have a positive impact on the model, those extending to the left have a negative impact on the model. Terms deemed significant for model projection are highlighted (blue). Analysis is based on two replicated arrays (Table 1). The dashed vertical lines indicate a term that is significant at the 0.10 level.

Table 1. Definitive screening design array. A definitive screening design was developed to assess the impact of 10 ingredients found within the chemically defined media. All concentrations are in g/L except trace elements, amino acids and vitamins which are ml / L. Trace element working concentration contained (g/L): 15 $FeSO_4.7H_2O$, 2.4 $MnSO_4.H_2O$, 2.4 $ZnSO_4.7H_2O$, and 0.48 $CuSO_4.5H_2O$. A 100 × stock amino acid mix contained (g/L): 12.9 arginine, and 10 each of histidine, leucine and methionine. A 1000 × vitamin stock contained (g/L): 0.1 pyridoxine, 0.02 folic acid, 0.05 each of thiamine, riboflavin, niacin, pantothenic acid and nicotinamide. Abbreviations: Fru., fructose, T.E. Trace element mixture; A.A., amino acid mixture; Vit., vitamin mixture. The DSD was performed in replicate.

Media	Fru.	NaH_2PO_4	Na_2HPO_4	K_2SO_4	$MgSO_4$	$CaCl_2$	NH_4Cl	T.E.	A.A.	Vit.	OD1	OD2
1	20.25	0.1	0.1	0.01	0.01	0.01	0.01	0.01	0.05	0.01	1.06	1.37
2	20.25	6	6	4.8	2.8	0.8	4.8	4.8	19	4.8	1.14	1.62
3	0.5	3.05	0.1	4.8	0.01	0.01	0.01	4.8	19	4.8	0.52	0.47
4	40	3.05	6	0.01	2.8	0.8	4.8	0.01	0.05	0.01	0.09	0.78
5	0.5	6	3.05	0.01	2.8	0.01	0.01	0.01	19	4.8	0.77	0.71
6	40	0.1	3.05	4.8	0.01	0.8	4.8	4.8	0.05	0.01	-	-
7	40	6	0.1	2.405	2.8	0.01	4.8	4.8	19	0.01	0.93	0.69
8	0.5	0.1	6	2.405	0.01	0.8	0.01	0.01	0.05	4.8	0.11	0.28
9	0.5	6	0.1	4.8	1.405	0.01	4.8	0.01	0.05	0.01	0.80	0.64
10	40	0.1	6	0.01	1.405	0.8	0.01	4.8	19	4.8	1.62	1.48
11	0.5	6	6	0.01	2.8	0.405	0.01	4.8	0.05	0.01	0.29	0.32
12	40	0.1	0.1	4.8	0.01	0.405	4.8	0.01	19	4.8	1.86	1.41
13	0.5	6	6	4.8	0.01	0.8	2.405	0.01	19	0.01	0.41	0.55
14	40	0.1	0.1	0.01	2.8	0.01	2.405	4.8	0.05	4.8	-	0.01
15	0.5	0.1	6	4.8	2.8	0.01	4.8	2.405	0.05	4.8	0.17	0.43
16	40	6	0.1	0.01	0.01	0.8	0.01	2.405	19	0.01	1.69	1.67
17	0.5	0.1	0.1	4.8	2.8	0.8	0.01	4.8	9.525	0.01	0.17	0.28
18	40	6	6	0.01	0.01	0.01	4.8	0.01	9.525	4.8	-	0.19
19	0.5	0.1	0.1	0.01	2.8	0.8	4.8	0.01	19	2.405	0.48	0.70
20	40	6	6	4.8	0.01	0.01	0.01	4.8	0.05	2.405	-	0.04
21	20.25	3.05	3.05	2.405	1.405	0.405	2.405	2.405	9.525	2.405	1.76	1.70
22	0.5	0.1	6	0.01	0.01	0.01	4.8	4.8	19	0.01	0.86	0.71
23	0.5	6	0.1	0.01	0.01	0.8	4.8	4.8	0.05	4.8	0.10	0.19
24	40	6	0.1	4.8	2.8	0.8	0.01	0.01	0.05	4.8	1.64	1.56
25	40	0.1	6	4.8	2.8	0.01	0.01	0.01	19	0.01	0.20	0.23

lowest value because there was some evidence of a negative impact from the 2-level factor analysis (Fig. 2).

Examination of the distributions for DSD2 revealed that the media compositions that generated the highest cell densities (OD_{600nm} >2.0) were associated with low trace element concentrations and high amino acid content (Fig. SI 2). We therefore probed the putative interaction between amino acids and trace elements using a medium that was formulated to permit robust cell growth over 72 h. It was observed that the absence of trace elements did not adversely affect growth of *C. necator* but that the absence of amino acids did (Fig. 4A). Interestingly, simultaneous exclusion of both amino acids and trace elements resulted in cell densities comparable with the control. We hypothesized that in the absence of one or more of the amino acids (methionine, histidine, leucine and/or arginine), one or more of the components of the trace elements ($CuSO_4$, $FeSO_4$, $MnSO_4$ and/or $ZnSO_4$) inhibits growth of *C. necator*. This was tested first by formulating media without amino acids and withdrawing each trace element in turn (Fig 4B). Under these conditions, media without $CuSO_4$ but containing other trace elements resulted in growth comparable to the control, while all three formulations that contained $CuSO_4$ had reduced cell densities. These observations support the hypothesis that $CuSO_4$ - in the absence of amino acids - inhibits *C. necator* growth. To determine which amino acid(s) interacts with $CuSO_4$, a series of experiments were performed in which each amino acid was excluded in formulations with and without $CuSO_4$. The first observation was that at high concentrations of amino acids the presence or absence of $CuSO_4$ did not affect growth (Fig 4C). The second observation was that at lower concentrations of amino acids growth was partially suppressed in the presence of $CuSO_4$ but that this was exacerbated in the absence of $CuSO_4$. $CuSO_4$ is therefore an important medium component and cannot simply be excluded from formulations. Similar growth responses were seen in experiments in which either methionine or leucine were excluded (Fig 4D and E). If arginine, or most notably histidine, were removed from the medium, then the presence of $CuSO_4$ impaired growth of *C. necator* (Fig 4F and G). For media lacking histidine this effect was also observable when all other amino acid levels were kept high (Fig. 4G). From this data we conclude two points: first, that histidine protects against the inhibitory effects of copper, but second, $CuSO_4$ is an essential component of the media whose absence retards growth when amino acid content is restricted. Both $CuSO_4$ and histidine concentrations must therefore be balanced for robust growth.

Final data augmentation. With a greater understanding of which components and concentrations are required to construct a medium supporting robust growth we re-investigated the roles that NaH_2PO_4, K_2SO_4, $MgSO_4$ and NH_4Cl have on the system. In the original DSD these factors were not identified as being statistically significant but those experiments were run under conditions in which key components (e.g. fructose and $CaCl_2$ (Fig. 3) were at settings that have since been identified as resulting in poor or unreliable growth responses. We therefore wished to re-evaluated these factors under conditions in which Na_2HPO_4, $CaCl_2$, trace elements and amino acid concentrations were not disruptive to cell growth (Table SI 4). Fructose was set at either 5 g/L or 25 g/L. The results indicated that these components did indeed affect growth rate when primary factors are not restricting growth (Fig. 5). Some of the suggestions from the original DSD were confirmed. Most notable was the observation that increasing NH_4Cl had a detrimental affect on culture cell density. NaH_2PO_4 and K_2SO_4 had some detrimental impact if concentrations were low, whilst $MgSO_4$ concentrations were not significant. These results were true at both high and low concentrations of fructose.

Finally, it is important to draw a distinction between what is statistically significant and what is essential. We observe little growth in formulae lacking $MgSO_4$ yet $MgSO_4$ is not deemed statistically significant. This is because under our experimental conditions, all concentrations of $MgSO_4$ tested are in excess and do not limit to growth. There is therefore scope to reduce $MgSO_4$ concentrations if this was desirable.

Modelling the media formula-growth response landscape. At this stage, 64 different variant formulations have been experimentally assessed in duplicate across three dif-

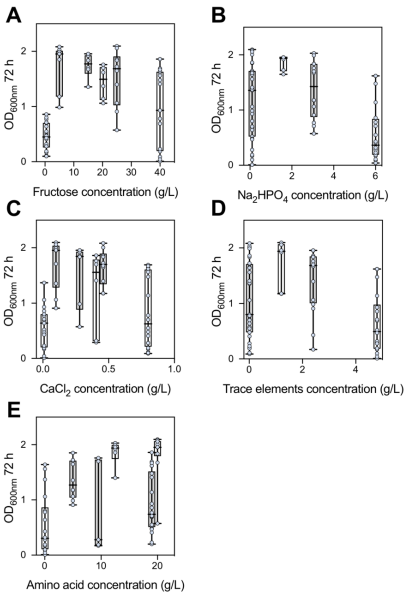


Fig. 3. Combined data for DSD1 and DSD2 for key media components. A. Fructose. **B.** Na2HPO4 **C.** CaCl2 **D.** Trace elements **E.** Amino acids.

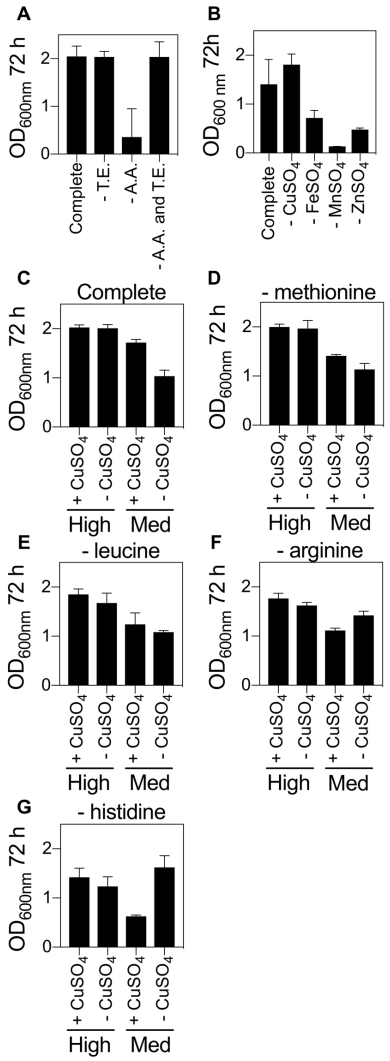


Fig. 4. Interactions between trace elements and amino acids. A. *C. necator* grown in a complete, chemically defined medium with either amino acids, trace elements or both amino acids and trace elements excluded. **B.** *C. necator* grown in a complete medium with each of the four trace elements excluded. **C-G** *C. necator* grown in a complete, chemically defined medium with or without $CuSO_4$ and with each of the four amino acids excluded. Error bars are S.E. Mean, $n = 2$.

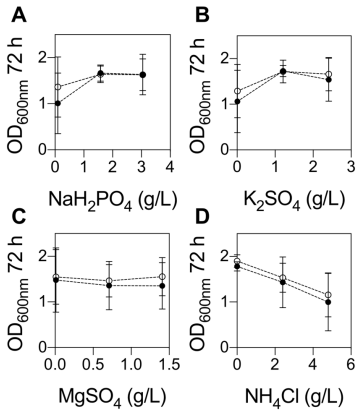


Fig. 5. Re-examination of non-significant media components. The impact of components not deemed statistically significant in DSD1 were re-examined under less extreme conditions. These were **A.** NaH_2PO_4 **B.** K_2SO_4 **C.** $MgSO_4$ and **D.** NH_4Cl. Experiments were conducted at 5 g/L (open circles) or 25 g/L (closed circles) fructose. Error bars are S.E. Mean, $n \geq 6$

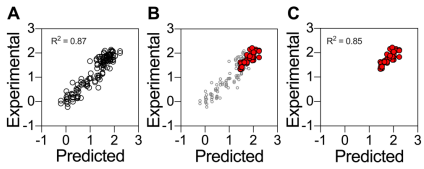


Fig. 6. Experimental validation of model predictions in 48-well plate format. A. Model predicted values plotted against experimental data for least squares model. **B.** New experimental data for 1 mL culture volumes overlaid against the original least squares model. Open circles show the training data set, red circles are new data. **C.** New experimental data for 1 mL cultures plotted on their own against the model predicted values. Three replicates were assessed for each prediction.

ferent experiments (25, 13 and 26 experimental runs respectively). We then built a statistical model, trained against this data set, that could describe our understanding of how the cell cultures respond to changes in media composition and predict performance in novel formulations. We performed two-level screening on all 128 runs. This identified a number of factors and factor interactions deemed significant for model projection. The screening did not highlight fructose or K_2SO_4 (as these had been at held concentrations that did not significantly impact growth during much of DSDS2 and DSD3) but these terms were included manually in the model as we knew they were important factors. These terms were used to construct a standard least squares model. The least squares model was able to describe the relationships within the data with good accuracy (Fig. 6A). Though the model was internally consistent it was important to know if it could predict OD_{600nm} at 72 h in formulations it had not encountered during model training. We assessed 16 new formulations with sampling biased towards media formulations that were predicted to be in the top 25% of media performance. Each of these was assessed in triplicate and the resulting OD_{600nm} compared to predicted OD_{600nm} (Fig. 6B,C). As predicted by the model, all of the new formulations fell within the upper quartile of formula performance.

The model allowed us to visualise phenomenon observed during the data-collection phase of the investigation. For example, it visualises the interaction between amino acids (specifically histidine) and trace elements (specifically copper) that was elucidated in Fig. 4. It shows that low and high concentrations of trace elements are detrimental to growth (low concentrations are to be avoided as these are essential to growth, while high concentrations may be toxic) and that increasing the amino acid concentration can help mitigate the inhibitory effects of high concentrations of trace elements (Fig. 7A). The greater the concentration of trace elements included the higher the amino acid concentration needs to be. Nevertheless, increasing the concentration of amino acids impacts in other areas. When fructose is low, increasing the concentration of amino acids increases cell density (Fig 7B). Given that the amino acids present in the media are the only alternative source of carbon for growth, raising the concentration of amino acids too high will impact on interpretation of experiments designed to examine carbon utilisation. A balance therefore needs to be struck between fructose, amino acids and trace element concentrations. The model also visualises interactions between Na_2HPO_4 and fructose (Fig. 7C). Greater concentrations of Na_2HPO_4 results in lower cell densities - an effect that can be partially offset by decreasing fructose concentrations. Finally, we had previously observed a negative effect of increasing NH_4Cl concentrations (Fig. 5D). The model indicates that this can be mitigated by increasing the K_2SO_4 concentration (Fig. 7D). Understanding the interactions between media components is vital for predictions of culture performance and allows the experimenter to alter media formulations for different experimental goals.

Model validation at greater volumes. We conducted two further model validation tests. We transferred these formulations to both 100 mL scale in shake flasks and 1 L batch fermentation in a bioreactor. Shake flask cultivations were carried out under identical conditions (30°C, 200 rpm) over 72 h periods in 250 mL baffled and non-baffled flasks, each containing 100 mL of medium. The formula for each cultivation was selected randomly from an L32 fractional factorial design of experiment (Fig SI 3). Growth (OD_{600nm}) for each formula in both types of flasks were strikingly similar at every interval throughout the cultivation period. Most importantly, the growth rank for baffled and non-baffled flasks were identical. Spearman's correlation showed a significant ($p < 0.05$) relationship between predicted and actual ranks for both flask types, with 0.867 correlation co-efficient ρ (Fig. 8A). Next, three formulae from shake flask cultivation together with two additional formulae were cultivated in 1 L bioreactor (Fig SI 4). Similarly, growth rank was predictable, with significant relationship between predicted rank and actual growth rank ($p = 0.900$) (Fig. 8B). During the culti-

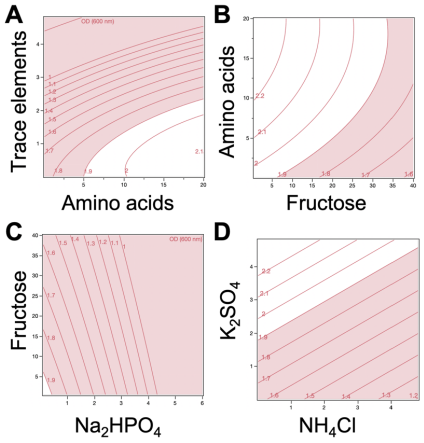


Fig. 7. Interactions of components of the defined media described by the least squares model. A. Amino acid and trace element interactions. **B.** Fructose and amino acid interactions. **C.** Disodium phosphate and fructose interactions. **D.** Ammonium chloride and potassium sulphate interactions. Each panel represents a two way-interaction. Red zones are areas where OD_{600nm} at 72 h fails to reach 1.9, and white zones where it surpasses 1.9. Each contour grid line represents an OD_{600nm} of 0.1. All other media components were kept at concentrations permitting high growth.

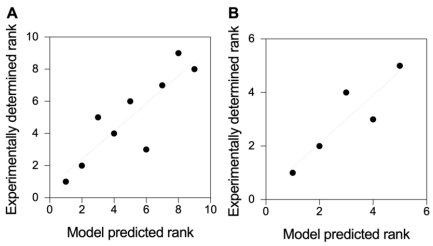


Fig. 8. Experimental validation of model predictions at shake-flask and bioreactor scale. A. Spearman's rank correlations between model predicted rank and experimental data for 100 mL culture volumes **B.** Spearman's rank correlations between model predicted rank and experimental data for 1 L bioreactor batch flask cultures.

vations, there were no significant changes in bioprocess parameters. Constant agitation at 200 rpm, with 1 vvm airflow rate was sufficient to maintain dissolved oxygen (dO_2) above 20%. Although no base was added in all cultivations, the pH of the media did not drop below 4.5, the set point. There was no significant relationship ($p = 0.233$; $p = 0.520$) between growth at 72 h and pH of media prior to inoculation or after growth.

Discussion

We developed a model trained against a structured data-set for the cultivation of *C. necator* H16 in a chemically defined medium with fructose as the primary carbon source. Our approach identified significant growth factors and their effects on culture density at 72 h. Fructose, glucose, glycerol and sucrose were used in the preliminary phase with fructose considered as the best carbon source supporting robust growth under heterotrophic conditions. While *C. necator* H16 has been reported to have broad substrate range its ability to utilize carbohydrates as a carbon source during heterotrophic growth appears to be limited to fructose and *N*-acetylglucosamine (8, 17, 29). Utilisation of fructose by *C. necator* H16 is most likely to occur via substrate import via an ATP-binding cassette (ABC-type) transporter, followed by catabolism via 2-keto-3-deoxy-6-phosphogluconate (KDPG) and the Entner-Doudoroff pathway. The responsible genes, notably a putative regulator (*frcR*), ribosome transporters (i.e. *frcA*, *frcC* and *frcB* orthologs in *Escherichia coli* and *Ralstonia solanacearum*) and other essential genes facilitating such metabolism are located on chromosome 2 inside gene clusters for glucose, 2-ketogluconate, and glucosamine catabolism (17). In contrast, phoshphofructokinase and 6-phosphogluconate dehydrogenase, key enzymes of the Embden-Meyerhoff-Parnas and oxidative pentose phosphate pathways, respectively, appear to be absent from the *C. necator* H16 genome. It is perhaps therefore not surprising that glucose supported little or no growth yet it has been reported that prolonged cultivation (> 70 h) with glucose as the sole carbon source resulted in a mutant that was able to utilize glucose (30). The mutant acquired such ability by mutating the *N*-acetylglucoasmine phosphotransferase system, which when deleted led to inability to utilize glucose (30, 31). Glycerol supported low growth of *C. necator* H16 especially in the medium forcing during the scoping trial. Such low growth was attributed to oxidative stress resulting from high levels of reactive oxygen species (ROS) formed as a result of elevated level of hydrogenases, leading to cell destruction by damaging DNA and other cell components (32).

Formulations of chemically defined media previously described for the cultivation of *C. necator* H16 are typically prepared without amino acids (3, 7, 11, 12, 22, 33). In this study, bacterial growth is improved when a small number of amino acids (arginine, histidine, leucine and methionine) are included in the medium. The data suggest that they serve as preferred sources of nitrogen, compared to NH_4Cl. Their inclusion resulted in a shorter lag phase compared to media lacking amino acids. Although the effect of histidine is greater than methionine, arginine and leucine, their action was synergistic. Interestingly, the model indicated that there was an interaction between fructose and amino acid content that indicates the carbon : nitrogen ratio balance must be maintained, irrespective of the actual values. Growth reductions are less under low amino acid and low fructose concentrations (5 ml/L and 5 g/L, respectively) compared to that under low amino acid and high fructose concentrations (5 ml/L and > 20 g/L, respectively).

Metal ions, especially divalent cations of d-block transition metals, are important for bacterial growth where they act as metalloenzyme cofactors in living cells. Their presence in high concentrations however, tends to be detrimental to cells (34). Copper is known to play diverse structural and catalytic

functions owing to its ability to exist either in a reduced (Cu^{+}) state with affinity for thiol and thioether groups, or an oxidized (Cu^{2+}) state with more likely coordination for oxygen or imidazole nitrogen group of amino acids including histidine (35). Bacteria have been reported to respond to high levels of copper by one of three families of metalloregulatory repressors: CopY, CsoR, and CueR. Under high concentration, an integral membrane Cu^{+} transporter (P1B-type ATPase) - characterised by histidine-rich domains, and conserved cytosine/histidine motifs within specific transmembrane domains - exports copper from the cytoplasm into the periplasm where in Gram-negative bacteria further detoxification and exportation are carried out by other related enzymes (35). Further, studies have reported that copper toxicity affected sugar (glucose) utilization in different microorganisms, and its growth inhibition on fungi (*Neurospora crassa* and *Saccharomyces cerevisiae*) was due to impaired/suppression of histidine biosynthesis (36, 37). In general, the toxic effect of copper appears to be enhanced under histidine limitation or impaired histidine biosynthesis, and is neutralized by addition of histidine.

This study provides insight into the impact of media formulations on growth and cell density of *C. necator* H16. Our growth model shows that the bacterium is able to grow to high cell density with minimal concentration of each growth component, provided the concentration of amino acids and trace elements are balanced. When cultivated under our formulae, growth peaked at 72 h and remained stationary after a further 20 h. Fructose, amino acids, and $CaCl_2$, and interactions between amino acids and fructose, fructose and $CaCl_2$, and amino acids and trace elements have significant positive effects on cell density, whilst Na_2HPO_4, and trace elements, and interactions between fructose and Na_2HPO_4, have negative effects. Addition of amino acids shortens the lag phase of growth and results in reproducible high cell density over repeated experiments. Besides fructose, only magnesium is considered essential for growth, whilst amino acids and K_2SO_4 are considered important. Thus, a defined medium supporting high cell density of *C. necator* H16 lacking some components (i.e. Na_2HPO_4, $CaCl_2$, NH_4Cl, and trace elements) can still result in high cell density comparable to formulae with all components. Both the information provided and the statistical approaches taken in this study will inform further efforts aimed at optimising other *C. necator* responses. This includes the biosynthesis of polyhydroxyalkanoate, platform chemicals, proteins and other products as well as future optimisation of growth under lithoautotrophic conditions with CO_2 and molecular hydrogen serving as carbon and energy sources, respectively.

Materials and Methods

Bacterial strains. *Cupriavidus necator* H16 (DSM 428) was obtained from Deutsche Sammlung von Mikrooganismen und Zellkulturen GmbH, German Collection of Microorganisms and Cell Cultures (DSMZ). The bacterial strain was resuscitated on nutrient agar (peptone 5 g/L, and meat extract 3 g/L) according to the supplier's instructions, and incubated at 30°C for 48 h.

Chemicals. Carbon sources (glucose, fructose, glycerol and sucrose), salts (with exception of $MgSO_4.H_2O$ and NH_4Cl), trace metals, amino acids (histidine, leucine and arginine) and some vitamins (thiamine, niacin, pantothenic acid) were obtained from Sigma-Aldrich. The remainder of the vitamins, $MgSO_4.H_2O$, NH_4Cl and methionine were obtained from Duchefa Biochemie B.V., BDH chemicals and Formedium, respectively.

Preparation of media. Solutions of glucose, fructose, sucrose, vitamins, amino acid and each trace metal solution were filter sterilized through 0.22 μm filter. Glycerol, NaH_2PO_4, Na_2HPO_4, $MgSO_4.H_2O$, NH_4Cl, K_2SO_4, $CaCl_2.2H_2O$ solutions were autoclaved. The trace solution was made of $CuSO_4.5H_2O$ (dissolved in 0.1M HCl), $FeSO_4.7H_2O$ (freshly prepared during each trace reconstitution from individual stock), $MnSO_4.H_2O$, and $ZnSO_4.7H_2O$. Amino acid stock solution contained: arginine, histidine, leucine, and methionine, while vitamin stock solutions contained: folic acid, niacin, nicotinamide, pantothenic acid, pyridoxine, riboflavin and thiamine. Subsequently, medium components were added from individual stock solutions except for trace elements which were added from reconstituted working solution. Forty-eight well plates were used in all trials, and medium reconstitution in each well was carried out using an automated liquid handling system (Eppendorf epMotion M5073). All stock solutions were prepared using water as a solvent, and were further diluted in sterile distilled water unless stated otherwise.

Inoculate preparation and growth measurement. For each experiment, 48 h colonies from LB agar were washed twice in sterile distilled water and diluted to a working inoculate concentration in the range of 108 cfu/ml. The inoculate was further diluted 1:100 in wells containing medium. Plates were incubated at 30°C, 170 rpm. Optical density (OD) at 600 nm was measured every 24 h using a Varioskan LUX™ Multimode Microplate reader (Thermo Scientific).

Batch cultivation in a bioreactor system. Large-scale cultivations were carried out in a batch mode with 1 L chemically defined media contained in 2 L capacity fermentors (Applikon ADI fermentation system). During the cultivations, pH was maintained above 4.5, temperature at 30°C, agitation at 200 rpm. Dissolved oxygen (dO_2) was maintained above 20 % (1, 6) with airflow at 1 vvm (volume of air per volume of medium). No anti-foam agents or base were added during cultivation. Starter culture media were of the same formulations with that used in fermenters, and were prepared by growing 100 mL culture in 250 mL non-baffled flasks to late exponential growth at 30°C, 200 rpm. Following polarization and calibration of the dO_2 probe, fermenters were inoculated with 10 mL (48 h) starter culture and cultivations were monitored on-line and off-line over 72 h. Samples were taken every 24 h for off-line OD_{600nm}.

Data analyses. Experimental designs were created using JMP Pro statistical software (version 13.0) and data from each experiment was analysed using the same software. Graphics were generated using GraphPad Prism 7.0.

ACKNOWLEDGEMENTS
Christopher C. Azubuike is a Commonwealth Scholar (NGCA-2016-060) funded by Department for International Development (DFID), UK government. Manuscript prepared using a modified version the HenriquesLab template available via www.overleaf.com.

Supporting Information

- **Table SI 1. Previously described chemically defined media for the growth of *C. necator*.** All concentrations are g/L unless otherwise indicated.
- **Table SI 2. Scoping Experiment.** A scoping trial was developed to assess the impact of 10 basic ingredients found within the chemically defined media. All concentrations are in g/L except trace elements, amino acids and vitamins which are ml / L. Abbreviations: T.E. Trace element mixture; A.A., amino acid mixture; Vit., vitamin mixture. Carbon: fructose, glucose, glycerol or sucrose. The medium trial was performed in duplicate..
- **Table SI 3. Definitive Screening Design 2.** A second definitive screening design was developed to assess the impact of the five ingredients identified as important to the chemically defined media. All concentrations are in g/L. Abbreviations with exception of T.E. and A.A, which are in ml/L: Fru., fructose, T.E., Trace element mixture; A.A., amino acid mixture. The DSD array was performed in duplicate.
- **Table SI 4. Definitive Screening Design 3.** A final set of experiments were performed to re-assess the impact of components not deemed significant in the first DSD. All concentrations are in g/L. Abbrevations: Fru., fructose, T.E. Trace element mixture; A.A., amino acid mixture. The DSD array was performed in duplicate.
- **Fig. SI 1. Growth characteristics of *C. necator* from initial scoping experiments. A.** Colony forming units against OD_{600nm}. Growth of *C. necator* on **B.** Glucose, **C.** Glycerol, and **D.** Sucrose as sole carbon source.
- **Fig. SI 2. Distribution of data for Definitive Screening Design 2.** Highlighted in blue cross-hatch are the settings that resulted in the greatest OD_{600nm} at 72 h for DSD2. Fructose, Na_2HPO_4 and $CaCl_2$ may be at either the highest or lowest concentrations, whereas trace elements and amino acid concentrations are found at the lowest or highest concentrations respectively.
- **Fig. SI 3. 100 mL validation experiment.** Nine different formulations were assessed at 100 mL culture volumes. For each row, the model predicted rank is shown on the left hand side, media settings for each component are shown in red underneath, the measured $OD_{(600nm)}$ 72 h for two replicate experiments and experimentally determined rank on the right hand side.
- **Fig. SI 4. 1 L validation experiment.** Five different formulations were assessed at 1 L culture volumes. For each row, the model predicted rank is shown on the left hand side, media settings for each component are shown in red underneath, the measured $OD_{(600nm)}$ 72 h for two replicate experiments and experimentally determined rank on the right hand side.

Bibliography

1. E. Grousseau, J. Lu, N. Gorret, S. E. Guillouet, and A. J. Sinskey. Isopropanol production with engineered cupriavidus necator as bioproduction platform. *Appl Microbiol Biotechnol*, 98(9):4277–90, 2014. ISSN 1432-0614 (Electronic) 0175-7598 (Linking). doi: 10.1007/s00253-014-5591-0.
2. M. Schmidt, J. L. Ienczak, L. K. Quines, K. Zanfonato, W. Schmidell, and G. M. F. de Aragao. Poly(3-hydroxybutyrate-co-3-hydroxyvalerate) production in a system with external cell recycle and limited nitrogen feeding during the production phase. *Biochemical Engineering Journal*, 112:130–135, 2016. ISSN 1369-703x. doi: 10.1016/j.bej.2016.04.013.
3. K. K. Aneja, R. D. Ashby, and D. K. Solaiman. Altered composition of *Ralstonia eutropha* poly(hydroxyalkanoate) through expression of pha synthase from *Allochromatium vinosum* atcc 35206. *Biotechnol Lett*, 31(10):1601–12, 2009. ISSN 1573-6776 (Electronic) 0141-5492 (Linking). doi: 10.1007/s10529-009-0052-z.
4. C. J. Brigham, C. F. Budde, J. W. Holder, Q. Zeng, A. E. Mahan, C. Rha, and A. J. Sinskey. Elucidation of beta-oxidation pathways in *Ralstonia eutropha* h16 by examination of global gene expression. *J Bacteriol*, 192(20):5454–64, 2010. ISSN 1098-5530 (Electronic) 0021-9193 (Linking). doi: 10.1128/JB.00493-10.
5. C. H. Bi, P. Su, J. Muller, Y. C. Yeh, S. R. Chhabra, H. R. Beller, S. W. Singer, and N. J. Hillson. Development of a broad-host synthetic biology toolbox for *Ralstonia eutropha* and its application to engineering hydrocarbon biofuel production. *Microbial Cell Factories*, 12, 2013. ISSN 1475-2859. doi: Artn10710.1186/1475-2859-12-107.
6. J. Marc, E. Grousseau, E. Lombard, A. J. Sinskey, N. Gorret, and S. E. Guillouet. Over expression of groesl in cupriavidus necator for heterotrophic and autotrophic isopropanol production. *Metab Eng*, 42:74–84, 2017. ISSN 1096-7184 (Electronic) 1096-7176 (Linking). doi: 10.1016/j.ymben.2017.05.007.
7. H Li, P H Opgenorth, D G Wernick, S Rogers, T Y Wu, W Higashide, P Malati, Y X Huo, K M Cho, and J C Liao. Integrated electromicrobial conversion of co2 to higher alcohols. *Science*, 335(6076):1596–1596, 2012.
8. R. Cramm. Genomic view of energy metabolism in ralstonia eutropha h16. *J Mol Microbiol Biotechnol*, 16(1-2):38–52, 2009. ISSN 1660-2412 (Electronic) 1464-1801 (Linking). doi: 10.1159/000142893.
9. S. Gruber, J. Hagen, H. Schwab, and P. Koefinger. Versatile and stable vectors for efficient gene expression in ralstonia eutropha h16. *J Biotechnol*, 186:74–82, 2014. ISSN 1873-4863 (Electronic) 0168-1656 (Linking). doi: 10.1016/j.jbiotec.2014.06.030.
10. J. Lu, C. J. Brigham, C. S. Gai, and A. J. Sinskey. Studies on the production of branched-chain alcohols in engineered ralstonia eutropha. *Appl Microbiol Biotechnol*, 96(1):283–97, 2012. ISSN 1432-0614 (Electronic) 0175-7598 (Linking). doi: 10.1007/s00253-012-4320-9.
11. T Oda, K Oda, H Yamamoto, A Matsuyama, M Ishii, Y Igarashi, and H Nishihara. Hydrogen-driven asymmetric reduction of hydroxyacetone to (r)-1,2-propanediol by ralstonia eutropha transformant expressing alcohol dehydrogenase from from kluyveromyces lactis. *Microbial Cell Factories*, 12, 2013.
12. J. Muller, D. MacEachran, H. Burd, N. Sathitsuksanoh, C. Bi, Y. C. Yeh, T. S. Lee, N. J. Hillson, S. R. Chhabra, S. W. Singer, and H. R. Beller. Engineering of ralstonia eutropha h16 for autotrophic and heterotrophic production of methyl ketones. *Appl Environ Microbiol*, 79(14):4433–9, 2013. ISSN 1098-5336 (Electronic) 0099-2240 (Linking). doi: 10.1128/AEM.00973-13.
13. J. S. Chen, B. Colon, B. Dusel, M. Ziesack, J. C. Way, and J. P. Torella. Production of fatty acids in ralstonia eutropha h16 by engineering beta-oxidation and carbon storage. *PeerJ*, 3:e1468, 2015. ISSN 2167-8359 (Print) 2167-8359 (Linking). doi: 10.7717/peerj.1468.
14. D. Przybylski, T. Rohwerder, C. Dilssner, T. Maskow, H. Harms, and R. H. Muller. Exploiting mixtures of h2, co2, and o2 for improved production of methacrylate precursor 2-hydroxyisobutyric acid by engineered cupriavidus necator strains. *Appl Microbiol Biotechnol*, 99(5):2131–45, 2015. ISSN 1432-0614 (Electronic) 0175-7598 (Linking). doi: 10.1007/s00253-014-6266-6.
15. L. Crepin, E. Lombard, and S. E. Guillouet. Metabolic engineering of cupriavidus necator for heterotrophic and autotrophic alka(e)ne production. *Metab Eng*, 37:92–101, 2016. ISSN 1096-7184 (Electronic) 1096-7176 (Linking). doi: 10.1016/j.ymben.2016.05.002.
16. G. C. Barnard, G. E. Henderson, S. Srinivasan, and T. U. Gemgross. High level recombinant protein expression in *Ralstonia eutropha* using t7 rna polymerase based amplification. *Protein Expr Purif*, 38(2):264–71, 2004. ISSN 1046-5928 (Print) 1046-5928 (Linking). doi: 10.1016/j.pep.2004.09.001.
17. A. Pohlmann, W. F. Fricke, F. Reinecke, B. Kusian, H. Liesegang, R. Cramm, T. Eitinger, C. Ewering, M. Potter, E. Schwartz, A. Strittmatter, I. Voss, G. Gottschalk, A. Steinbuchel, B. Friedrich, and B. Bowien. Genome sequence of the bioplastic-producing "knallgas" bacterium ralstonia eutropha h16. *Nat Biotechnol*, 24(10):1257–62, 2006. ISSN 1087-0156 (Print) 1087-0156 (Linking). doi: 10.1038/nbt1244.
18. A. Lykidis, D. Perez-Pantoja, T. Ledger, K. Mavromatis, I. J. Anderson, N. N. Ivanova, S. D. Hooper, A. Lapidus, S. Lucas, B. Gonzalez, and N. C. Kyrpides. The complete multipartite genome sequence of cupriavidus necator jmp134, a versatile pollutant degrader. *PLoS*

One, 5(3):e9729, 2010. ISSN 1932-6203 (Electronic) 1932-6203 (Linking). doi: 10.1371/journal.pone.0009729.

19. N. Berezina, B. Yada, and R. Lefebvre. From organic pollutants to bioplastics: insights into the bioremediation of aromatic compounds by *Cupriavidus necator*. *N Biotechnol*, 32(1):47–53, 2015. ISSN 1876-4347 (Electronic) 1871-6784 (Linking). doi: 10.1016/j.nbt.2014.09.003.
20. O. P. Peoples and A. J. Sinskey. Poly-beta-hydroxybutyrate (phb) biosynthesis in alcaligenes eutrophus h16. identification and characterization of the phb polymerase gene (phbc). *J Biol Chem*, 264(26):15298–303, 1989. ISSN 0021-9258 (Print) 0021-9258 (Linking).
21. S. Obruca, I. Marova, M. Stankova, L. Mravcova, and Z. Svoboda. Effect of ethanol and hydrogen peroxide on poly(3-hydroxybutyrate) biosynthetic pathway in cupriavidus necator h16. *World J Microbiol Biotechnol*, 26(7):1261–7, 2010. ISSN 0959-3993 (Print) 0959-3993 (Linking). doi: 10.1007/s11274-009-0296-8.
22. P. K. Sharma, J. Fu, V. Spicer, O. V. Krokhin, N. Cicek, R. Sparling, and D. B. Levin. Global changes in the proteome of cupriavidus necator h16 during poly-(3-hydroxybutyrate) synthesis from various biodiesel by-product substrates. *AMB Express*, 6(1):36, 2016. ISSN 2191-0855 (Print) 2191-0855 (Linking). doi: 10.1186/s13568-016-0206-z.
23. Kedar G. Patel, Mark Welch, and Claes Gustafsson. *Leveraging gene synthesis, advanced cloning techniques, and machine learning for metabolic pathway engineering*, book section Chapter 4, pages 53–71. Springer International Publishing, 2016. ISBN 978-3-319-41964-0 978-3-319-41966-4. doi: 10.1007/978-3-319-41966-4_4.
24. P. Xu, E. A. Rizzoni, S. Y. Sul, and G. Stephanopoulos. Improving metabolic pathway efficiency by statistical model-based multivariate regulatory metabolic engineering. *ACS Synth Biol*, 6(1):148–158, 2017. ISSN 2161-5063 (Electronic) 2161-5063 (Linking). doi: 10.1021/acssynbio.6b00187.
25. F. Caschera, M. A. Bedau, A. Buchanan, D. Gazzola G. Hanczyc M. Cawse, J. de Lucrezia, and N. H. Packard. Coping with complexity: machine learning optimization of cell-free protein synthesis. *Biotechnol. Bioeng.*, 108(9):2218–2228, 2011.
26. C. Gustafsson, J. Minshull, S. Govindarajan, J. Ness, A. Villalobos, and M. Welch. Engineering genes for predictable protein expression. *Protein Expression and Purification*, 83 (1):37–46, 2012. ISSN 1046-5928.
27. S. R. Brown, M. Staff, R. Lee, J. Love, D. A. Parker, S. J. Aves, and T. P. Howard. Design of experiments methodology to build a multifactorial statistical model describing the metabolic interactions of alcohol dehydrogenase isozymes in the ethanol biosynthetic pathway of the yeast saccharomyces cerevisiae. *ACS Synth Biol*, 2018. ISSN 2161-5063 (Electronic) 2161-5063 (Linking). doi: 10.1021/acssynbio.8b00112.
28. S. Govindarajan, B. Mannervik, J. A. Silverman, K. Wright, D. Regitsky, U. Hegazy, T. J. Purcell, M. Welch, J. Minshull, and C. Gustafsson. Mapping of amino acid substitutions conferring herbicide resistance in wheat glutathione transferase. *ACS Synth Biol*, 4(3):221–7, 2015. ISSN 2161-5063 (Electronic) 2161-5063 (Linking). doi: 10.1021/sb500242x.
29. S. Sichwart, S. Hetzler, D. Broker, and A. Steinbuchel. Extension of the substrate utilization range of ralstonia eutropha strain h16 by metabolic engineering to include mannose and glucose. *Appl Environ Microbiol*, 77(4):1325–34, 2011. ISSN 1098-5336 (Electronic) 0099-2240 (Linking). doi: 10.1128/AEM.01977-10.
30. A. Franz, R. Rehner, A. Kienle, and H. Grammel. Rapid selection of glucose-utilizing variants of the polyhydroxyalkanoate producer ralstonia eutropha h16 by incubation with high substrate levels. *Lett Appl Microbiol*, 54(1):45–51, 2012. ISSN 1472-765X (Electronic) 0266-8254 (Linking). doi: 10.1111/j.1472-765X.2011.03171.x.
31. M. Raberg, J. Bechmann, U. Brandt, J. Schluter, B. Uischner, B. Voigt, M. Hecker, and A. Steinbuchel. Versatile metabolic adaptations of ralstonia eutropha h16 to a loss of pdhl, the e3 component of the pyruvate dehydrogenase complex. *Appl Environ Microbiol*, 77 (7):2254–63, 2011. ISSN 1098-5336 (Electronic) 0099-2240 (Linking). doi: 10.1128/AEM.02360-10.
32. E. Schwartz, B. Voigt, D. Zuhlke, A. Pohlmann, O. Lenz, D. Albrecht, A. Schwarze, Y. Kohlmann, C. Krause, M. Hecker, and B. Friedrich. A proteomic view of the facultatively chemolithoautotrophic lifestyle of ralstonia eutropha h16. *Proteomics*, 9(22):5132–42, 2009. ISSN 1615-9861 (Electronic) 1615-9853 (Linking). doi: 10.1002/pmic.200900333.
33. J. Lu, C. J. Brigham, C. Rha, and A. J. Sinskey. Characterization of an extracellular lipase and its chaperone from ralstonia eutropha h16. *Appl Microbiol Biotechnol*, 97(6):2443–54, 2013. ISSN 1432-0614 (Electronic) 0175-7598 (Linking). doi: 10.1007/s00253-012-4115-z.
34. P. Chandrangsu, C. Rensing, and J. D. Helmann. Metal homeostasis and resistance in bacteria. *Nat Rev Microbiol*, 15(6):338–350, 2017. ISSN 1740-1534 (Electronic) 1740-1526 (Linking). doi: 10.1038/nrmicro.2017.15.
35. J. M. Arguello, D. Raimunda, and T. Padilla-Benavides. Mechanisms of copper homeostasis in bacteria. *Front Cell Infect Microbiol*, 3:73, 2013. ISSN 2235-2988 (Print) 2235-2988 (Linking). doi: 10.3389/fcimb.2013.00073.
36. C. Subramanyam and G. Venkateswerlu. The effect of copper on histidine biosynthesis in neurospora crassa. *Journal of Biosciences*, 1(2):143–149, 1979. ISSN 0250-5991 0973-7138. doi: 10.1007/bf02706326.
37. D. A. Pearce and F. Sherman. Toxicity of copper, cobalt, and nickel salts is dependent on histidine metabolism in the yeast saccharomyces cerevisiae. *J Bacteriol*, 181(16):4774–9, 1999. ISSN 0021-9193 (Print) 0021-9193 (Linking).